ISSN:
(Online) 2978-4530
(Print) 2978-4522

Journal of Business and Digital Innovation

JBDI

Northumbria
UNIVERSITY

Vol. 1
January 2026
Issue 1 (Maiden Volume/Issue)

Journal of Business and Digital Innovation (JBDI)

Northumbria University, United Kingdom

Vol. 1, Issue 1 (Maiden Volume/Issue), January 2026. ISSN: (Online) 2978-4530, (Print) 2978-4522

Description

The Journal of Business and Digital Innovation (JBDI) explores the dynamic interplay between digital technologies and business practices, focusing on how digital transformation influences innovation, entrepreneurship, organisational performance, leadership and management capabilities. The journal seeks interdisciplinary research articles with quantitative, qualitative, mixed methods approaches and systematic literature review that contribute to the discourse on the implications of the prevailing industrial revolution for businesses and society. Currently, JBDI is published both online and through print media.

Aims and Scope

The Journal aims to advance the understanding of interaction between digital innovation and business management practices. The journal is inclined to publish insightful articles, and its scope encompasses a wide range of topics but not exhaustive of the following: *Digital Innovation in Leading and Managing Business; Human Resource Management and Digital Innovation; Business and Management of Digital Innovation; Strategic Marketing and Digital Innovation; Digital Innovation in Hospitality and Tourism Business; Trends in Business Analytics and Digital Marketing; Digital Innovation in Financial Management; Advancements of Digital Innovation in Entrepreneurship; Digital Innovation and Corporate Governance; Digital Innovation in International Project Management; Cyber Security for Business Performance and the Business of Cyber Security; Computing, Artificial Intelligence, and International Business; Ethical and Social Implications of Digital Innovation in Business; Digital Innovation and Sustainability Agenda; Digital Innovation in Global Logistics Operations and Supply Chain Management; Reviews on Theoretical Foundations of Digital Innovation in Business.*

Papers in this Issue

Editors' Note

It is with great enthusiasm and a deep sense of purpose that I welcome readers to the inaugural volume and first issue of the *Journal of Business and Digital Innovation (JBDI)*. The launch of this journal marks an important milestone in our collective effort to understand, interpret, and shape the rapidly evolving intersection between business practice and digital transformation. We are living in a period defined by unprecedented technological acceleration. Artificial intelligence, data analytics, digital platforms, automation, and new organisational models are reshaping industries, redefining competitive advantage, and challenging long-standing assumptions about how value is created and delivered. In this dynamic environment, scholarship must not only keep pace, but it must also lead with insight, clarity, and critical reflection. JBDI emerges as a response to this imperative with the mission is to provide a rigorous, interdisciplinary, and globally relevant platform for research that advances knowledge, explores the dynamic interplay between digital technologies and business practices in the digital age.

This maiden issue brings together contributions that reflect the diversity and depth of contemporary inquiry in the field. The articles featured here explore themes ranging from digital strategy and innovation management to emerging technologies, organisational transformation, and the socio-economic implications of digitalisation. Each piece has undergone a robust peer-review process to ensure the highest academic standards. As we embark on this journey, I extend my sincere appreciation to the authors who entrusted us with their work, the reviewers whose expertise strengthened each manuscript, and the editorial board whose commitment has shaped the vision of this journal. I am equally grateful to our readers, scholars, practitioners, policymakers, and students, whose engagement will give life and relevance to this publication.

The emergence of JBDI is not merely the beginning of a journal; it is the beginning of a community. A community committed to intellectual curiosity, methodological rigour, and meaningful impact. We invite you to join us in this endeavour by reading, contributing, critiquing, and collaborating. With this first issue, we set the foundation for what we hope will become a leading voice in the global discourse on business and digital innovation. We look forward to the ideas, debates, and discoveries that will emerge in the volumes ahead.

"AI-Driven Personalized Learning: Predicting Academic Performance Through Leadership Personality Traits" by **Nitsa J Herzog, Rejwan Bin Sulaiman, David J Herzog & Rose Fong** is the first article. This study demonstrated how AI-driven modelling and leadership personality traits can enhance personalized learning. Using data from 129 master's students and multiple self-assessment tools, the author showed that machine learning, especially Random Forest can reliably predict academic performance. The findings highlight the value of early identification of students' strengths and challenges, offering educators a powerful evidence-based pathway for tailoring support, improving learning outcomes, and advancing data-informed educational practice.

The second article is on *"Relationships between Airline Sustainability and Consumer Behaviour: An assessment of the influence of environmental awareness on the decision-making process of European airline customers"* by **Yves Kremer & Eustathios Sainidis** offered timely insight into how rising environmental awareness is reshaping passenger behaviour in Europe's aviation sector. While travellers increasingly value sustainability and support investments in greener fuels and carbon-offsetting, they remain highly price-sensitive. The findings reveal a widening gap between consumer expectations and the industry's perceived urgency on climate action. By highlighting these tensions, the research underscores the need for more credible, visible sustainability commitments across the airline industry.

"Preserving Cognitive Ownership of Academic Writing in Higher Education: A Sustainable Hybrid Pedagogical Framework for Reasoning-Centred Artificial Intelligence Integration" by **Sunika Naz; Stella Sardar & Imad Yasir Nawaz** is the third article. The authors examined how different generative AI practices such as augmentation, co-construction, and replacement affect students' reasoning and cognitive ownership in higher education writing. Using mixed-methods data from UK students, the study shows that augmentation enhances reflective reasoning, co-construction offers mixed outcomes, and replacement weakens autonomy. It introduces the Hybrid Human–AI Reasoning Integrity Model (HHARIM), a framework promoting ethical, human-centred AI use to protect learning integrity and support sustainable academic development.

The fourth article *"Influence of Personal Branding on Entrepreneurial Success of Fitness Coaches in The Uk"* by **Michelle Romero, Olugbenga Akintola, Emmanuel Nwachukwu & Mosunmola Adeyeye**. This article offers valuable insight into how personal branding shapes entrepreneurial success among UK fitness coaches. Survey evidence shows authenticity is the strongest individual predictor of success, while credibility is suggestive and attractiveness has no meaningful effect. However, an integrated branding approach combining all three dimensions significantly boosts outcomes. The findings provide practical guidance for coaches and highlight the need for further qualitative or mixed-method research to deepen understanding of branding's entrepreneurial impact.

In the fifth article, *"Risk in Outsourced Information Technology (IT) Operations: A Systematic Literature Review of Technological Uncertainty, Knowledge Management and Opportunistic Behaviour",* **Peter Hewes, Marcelo Martins Sa,**

Adrian Small & Dongjun Li conducted a systematic review of 63 studies examining technological uncertainty and risk in outsourced IT operations. It highlights how digital transformation drives efficiency but introduces vulnerabilities such as cyber-risk and information asymmetry. The review shows that knowledge management is a key mitigation mechanism and identifies major gaps, particularly limited attention to the ex-post operational phase. The study provides an integrated understanding of technology risk and outlines priorities for future sector-specific research.

Promise Akwaowo & Imani Silver Kyaruzi in the sixth article *"The Effect of Mobile Payment Methods on Customer Decisions on Jumia's Shopping Platform in Nigeria"* investigated how mobile payment technologies influence customer buying behaviour on Jumia in Nigeria. Drawing on survey data from 150 users, the study shows that customer trust strongly predicts mobile payment adoption and conversion, while infrastructural challenges exert minimal direct impact. The findings underscore trust and perceived security as central to successful digital transactions and offer practical implications for strengthening digital trust and optimising payment systems in Nigeria's growing mobile commerce landscape.

"Comparative Analysis of Traditional Machine Learning, Deep Learning, and Hybrid Ensemble Models for Anomaly Detection and Web Application Firewall Optimisation" by **Shiva Nezamzadeh & Dilek Celik** in the seventh article. This study offers a timely contribution to cybersecurity by comparing machine learning models for anomaly detection in web application firewalls. Through consistent preprocessing and class balancing, the authors demonstrate the strong performance of a CNN model, which rivals a more complex stacking ensemble while remaining efficient and interpretable. Their findings underscore the value of robust preprocessing and highlight CNNs as practical candidates for real-time security applications.

Banjo Hassan, Victor Adeyeye, Hassanat Hassan & Kehinde Shosanya in the eighth article *"Mission Statement Attributes and Employee Engagement in the Nigerian Banking Sector: Evidence from Ogun state, Nigeria"* underscored a powerful yet often underestimated truth: mission statements matter. By revealing how clarity, effective communication, and personal resonance of a mission significantly elevate employee engagement in Ogun State's banking sector, the research offers timely insight for leaders. Its evidence reminds organizations that engagement is not accidental. It grows when employees understand, connect with, and internalize a shared purpose that guides daily work and long-term direction.

"Sentiment Analysis of Public Perceptions on ChatGPT and Generative Artificial Intelligence (GenAI): Model's Performance Evaluation and Examining Benefits and Risks in Education and Healthcare" by **Mifta Uddin Khan & Dilek Celik** is the nineth article. This study investigated a timely, data-driven window into how society perceives GenAI as it becomes woven into everyday life. By analysing over a million posts, the authors reveal largely positive sentiment toward ChatGPT while highlighting persistent concerns across education and healthcare. Their rigorous comparison of ML, DL, and transformer models where BERT excels that strengthens the evidence base. The insights presented here will support more responsible, informed, and sector-sensitive AI adoption.

Welcome to Volume 1, Issue 1.
Welcome to the Journal of Business and Digital Innovation.

Editor-in-Chief: ***Eustathios Sainidis***
Journal of Business and Digital Innovation

Deputy Editor-in-Chief: ***Olugbenga Akintola***
Journal of Business and Digital Innovation

Table of Contents

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 1-16
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1741

AI-Driven Personalized Learning: Predicting Academic Performance Through Leadership Personality Traits

[1]Nitsa J Herzog, [2]Rejwan Bin Sulaiman, [3]David J Herzog & [4]Rose Fong
[1,2 & 4]Northumbria University, London
[3]Ulster University, London

Abstract
The study explores the potential of AI technologies in personalized learning, suggesting the prediction of academic success through leadership personality traits and machine learning modelling. The primary data were obtained from 129 master's students in the Environmental Engineering Department, who underwent five leadership personality tests with 23 characteristics. Students used self-assessment tools that included Personality Insight, Workplace Culture, Motivation at Work, Management Skills, and Emotion Control tests. The test results were combined with the average grade obtained from academic reports. The study employed exploratory data analysis and correlation analysis. Feature selection utilized Pearson correlation coefficients of personality traits. The average grades were separated into three categories: "Fail", "Pass", and "Excellent". The modelling process was performed by tuning seven ML algorithms, such as Support Vector Machine, Logistic Regression, K-Nearest Neighbors, Decision Tree, Gradient Boosting, Random Forest, XGBoost and LightGBM. The highest predictive performance was achieved with the Random Forest classifier, which yielded an accuracy of 87.50% for the model incorporating 17 personality trait features and the leadership mark feature, and an accuracy of 85.71% for the model excluding this feature. In this way, the study offers an additional opportunity to identify students' strengths and weaknesses at an early stage of their education process and select the most suitable strategies for personalized learning.

Keywords: AI for personalized learning; AI in higher education; Leadership personality traits; Performance modelling; grade prediction.

Wordcount: 211

1.0 Introduction

The popularity of customized education services and personalized learning/teaching approaches has attracted considerable attention in recent years. AI-driven learning is going to replace traditional methods (Luckin and Cukurova, 2019). The prediction of outcomes from the usage of AI technologies and their effects on the educational sector are actively evaluated (Ouyang and Zhang, 2024). Personalized learning heavily depends on the individual's character traits that reflect specific behavioural patterns (Vorobyeva et al., 2025). Understanding the individual traits that impact academic performance can significantly enhance students' learning experiences, increase their motivation, and improve their engagement in learning activities.

The modern world relies heavily on digitalization and teamwork. Digital leaders can be trained according to organizational needs. Many educational bodies include leadership training in their curriculum. Personal leadership traits can be identified through specialized career tests and strengthened using personal development strategies and approaches. Previous research evaluating the effect of personality types on academic achievement found strong correlations between them. However, AI technologies have opened up additional opportunities to utilize this knowledge in personalized teaching and learning processes (Halkiopoulos and Gkintoni, 2024). The motivation for this study was to investigate the dependencies between leadership traits, which were not explored before, and student performance by evaluating a comprehensive set of self-assessment results and creating an AI-driven predictive model.

The study aimed to answer several research questions:
1. How do leadership personality traits correlate with students' academic performance?

2. Which leadership personality traits appear most influential in predicting higher student grades?
3. Can a machine learning model accurately predict students' grade categories based on leadership personality traits?
4. How can universities use insights from this study to develop personalized academic strategies tailored to students' strengths and personality profiles?

The research was focused on completing the following objectives:
1. Identify a diverse group of students from different programs in the Environmental Engineering (EE) department.
2. Explore results from self-assessment, including Personality Insight test, Workplace Culture, Motivation at Work, Emotional Control, and Management Skills.
3. Align test results with students' average academic mark across several modules.
4. Analyze the dataset using quantitative statistical methods.
5. Determine which personality traits appear to be the most influential for academic success.
6. Create a machine learning model that can predict grade categories, such as "Pass", "Fail", and "Distinction".
7. Formulate recommendations for the university to enhance student outcomes and help to develop personalized academic strategies based on individual strengths and personality profiles.

The article consists of 5 sections. Section 2, "Related Works", represents the determinant of academic performance in the light of modern leadership theories, discusses the previous studies that explored personality types and students' achievements, and highlights AI-based studies targeting the predictions of students' grades. Section 3, "Materials and Methods", describes the data collection procedures and discusses the study's design. The following "Results" section focuses on data analytical parts, feature selection and modelling stages. The "Discussion and Conclusion" section discusses the teaching strategies towards personalized teaching and learning using ML predictions.

2.0. Related Works

This literature review explores modern leadership trait theories, determinants of student academic success, the role of personality in academic achievement, self-assessment as a measurement tool, and the latest advancements in AI for grade prediction and personality-based performance prediction. At the end of the literature review, a research gap is identified in integrating AI-based personality assessments with leadership theories to predict student performance in educational contexts.

2.1. The Modern Leadership Trait Theories Overview

In the early 20th century, as leadership theories began to emerge, researchers believed that leadership was an innate quality that leaders were born, not made. This belief was formalized in the Great Man Theory by Carlyle in his book "On Heroes, Hero-Worship and the Heroic in History" published in 1894 (Carlyle, 1993). The theory was supported by examples of leaders from privileged backgrounds, including those with elite education and wealthy families. However, following the Industrial and later the Technological Revolutions, the theory's relevance declined. It was later re-evaluated by Cawthon in "Leadership: The Great Man Theory Revisited" in 1996 (Organ, 1996).

Because the Great Man Theory lacked empirical support, a new approach emerged—the Trait Theory of Leadership. While it shared the idea that certain individuals possess qualities that make them more effective leaders, it differed in two keyways: it identified specific traits and skills that contribute to leadership success, and it proposed that these attributes can be developed through education and experience. Subsequently, researchers shifted their focus to behavioural and situational theories, which emphasized the influence of personal traits in relation to specific contexts (Germain, 2008).

In fact, Kirkpatrick and Locke in 1991 identified six traits (Kirkpatick and Locke, 1991) that distinguish effective leaders from others: drive (ambition and energy), desire to lead (motivational inclination toward leadership roles), honesty and integrity (trustworthiness and ethics), self-confidence (confidence in one's abilities), cognitive ability (intelligence and analytical skills), knowledge of the field (expertise in the relevant domain).

Modern leadership trait theory has evolved significantly from early "great man" theories to more nuanced, empirically supported frameworks. Contemporary research acknowledges that effective leadership stems from a complex interplay of traits, behaviours, and situational factors (Northouse, 2021). The Big Five personality model has become particularly influential in leadership research, with meta-analyses demonstrating consistent relationships between personality dimensions and leadership effectiveness (Judge, Colbert and Ilies, 2004). Current literature identifies several key trait categories that consistently correlate with leadership effectiveness:

• *Cognitive Abilities:* Intelligence, particularly emotional intelligence, has emerged as a critical leadership trait. Goleman's (Goleman, 1995). emotional intelligence framework encompasses self-awareness, self-regulation, motivation, empathy, and social skills. Meta-analytic evidence supports moderate to strong correlations between cognitive abilities and leadership performance (Judge et al., 2002).

• *Personality Dimensions:* The Five-Factor Model provides a robust framework for understanding leadership-relevant personality traits. Extraversion shows the strongest correlation with leadership emergence and effectiveness, followed by conscientiousness and openness to experience (Judge et al., 2002). Neuroticism typically shows negative correlations with leadership effectiveness.

• *Motivational Factors:* Achievement motivation, power motivation, and leadership motivation have been identified as crucial drivers of leadership behaviour. McClelland's research (McClelland, 1985) on implicit motives demonstrates that effective leaders often exhibit high power motivation combined with impulse control.

Recent developments in leadership trait theory include the integration of character strengths (Peterson and Seligman, 2004), authentic leadership traits (Avolio and Gardner, 2005), and servant leadership characteristics (Van Dierendonck, 2011). These approaches emphasize moral dimensions and follower development as core leadership competencies.

2.2. Review of Established Determinants/Academic Performance Factors of Student Success

A complex interplay of cognitive, non-cognitive, and environmental factors influences academic performance. Cognitive factors, particularly academic ability and intelligence are widely recognized as primary predictors of success in educational settings. Richardson et al.'s (Richardson, Abraham and Bond, 2012) meta-analysis demonstrated that prior academic performance is the strongest correlation of future GPA, with a correlation coefficient of 0.34. General intelligence and specific academic competencies consistently predict outcomes across all levels of education. However, non-cognitive traits also play a crucial role. Among personality dimensions, conscientiousness has emerged as a reliable predictor of academic success, with Poropat's (Poropat, 2014) meta-analysis showing its effect sizes rival those of intelligence. Additionally, self-regulation and effective study strategies, as outlined in Zimmerman's theory (Zimmerman, (2002), enhance academic outcomes by promoting goal setting and strategic learning behaviours. Motivation, particularly mastery goal orientation as described by Dweck (Dweck, 2006), is another key determinant, leading to sustained academic achievement. Environmental and social variables further influence educational attainment. Socioeconomic status impacts academic success through access to resources and cultural capital (Sirin, 2005) while social support systems, including family, peers, and faculty, have been shown to improve retention and performance, as emphasized in Tinto's model (Tinto, 1993). Institutional characteristics such as teaching quality and support services also contribute meaningfully to student outcomes (Pascarella, and Terenzini, 2005). Together, these factors underscore the multifaceted nature of academic achievement and emphasize the importance of considering both individual traits and contextual influences.

2.3. Previous Studies on Personality and Academic Achievement

Personality traits, especially conscientiousness, are strong predictors of academic performance, with consistent positive correlations due to factors like time management and persistence (Trapmann et al., 2007). Openness to experience also supports academic success in intellectually demanding contexts (Poropat, 2009), while neuroticism tends to hinder performance due to anxiety (Chamorro-Premuzic and Furnham, 2003). Beyond the Big Five, academic self-efficacy significantly predicts performance and persistence (Bandura, 1997), whereas the role of grit is debated, with limited value beyond conscientiousness (Duckworth et al., 2007). Cultural and contextual differences influence how personality traits relate to academic outcomes, varying across cultures, disciplines, and assessment types (Poropat, 2009).

2.4. Self-Assessment as a Measurement Tool and Its Reliability and Validity in Educational Contexts

Self-assessment, grounded in metacognitive theory, plays a central role in self-regulated learning by enabling students to evaluate their learning and performance (Zimmerman, 2002). Its effectiveness relies heavily on metacognitive awareness and the ability to accurately calibrate self-perceptions. In terms of reliability, self-report personality measures typically demonstrate good internal consistency, with Cronbach's alpha values ranging from 0.70 to 0.90 in tools like the NEO-PI-R and the Big Five Inventory (John and Srivastava, 1999). These assessments also show moderate to high test-retest reliability, particularly for core personality traits, with stability increasing with age and correlations above 0.70 over several years (Roberts and DelVecchio, 2000). Furthermore, the inter-rater agreement between self-reports and observer ratings typically falls between 0.40 and 0.60, indicating moderate convergent validity and the dis-tinct perspective that self-assessments offer (Connolly, Kavanagh and Viswesvaran, 2007).

However, self-assessment is limited by several validity concerns. One major issue is social desirability bias, where individuals present themselves in a more favourable light, thereby reducing the accuracy of their responses, particularly for traits like conscientiousness (Paulhus, 2002). Additionally, people often engage in self-enhancement, overestimating their abilities and traits, although some self-enhancement can positively relate to psychological well-being and performance (Taylor and Brown, 1988). Accuracy also varies by domain; students tend to be more precise in assessing their performance on concrete tasks than on complex or ambiguous ones (Dunning, Heath and Suls, 2004).

To improve the validity of self-assessment, training and scaffolding strategies, such as structured rubrics, exemplars, and guided practice, have proven effective in enhancing accuracy (Andrade, and Du, 2007). Technology can further support this process by offering immediate feedback and calibrated comparisons to objective measures. Digital platforms and adaptive systems help refine self-assessment accuracy by aligning subjective evaluations with actual performance outcomes (Panadero, Jonsson, and Botella, 2017).

2.5. AI in Education

Advancements in artificial intelligence and machine learning have significantly enhanced educational data mining and the prediction of student performance. Machine learning algorithms can process large-scale datasets comprising student demographics, behavioural patterns, and learning analytics to identify trends and predict academic outcomes with high accuracy. Supervised learning models such as random forests, support vector machines, and gradient boosting techniques outperform traditional statistical methods by capturing complex, non-linear interactions among variables (Romero, and Ventura, 2020). Deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), further enhance prediction accuracy by analysing sequential learning data and identifying temporal patterns in student behaviour (Ujkani, Minkovska and Hinov, 2024).

In parallel, learning analytics combined with big data methods has become central to performance forecasting. Clickstream data collected from learning management systems—tracking time on task, navigation behaviour, and interaction frequency—provides valuable insights when analysed through AI algorithms, allowing for the early identification of at-risk students using explainable AI (Chen et al., 2020). Integrating multimodal data, such as academic performance, demographic information, social networks, and behavioural signals, leads to more accurate and comprehensive predictive models (Dutt, Ismail and Herawan, 2017).

AI also facilitates real-time prediction and intervention through early warning systems that detect students at risk of failure or dropout, often achieving over 80% prediction accuracy, allowing for continuous monitoring throughout the term (Helal et al., 2018). Adaptive learning platforms, driven by intelligent tutoring systems, utilize predictive analytics to tailor learning content to individual student profiles, dynamically adjusting the difficulty and pacing to optimize learning outcomes (Chrysafiadi, and Virvou, 2013).

Natural language processing (NLP) has further expanded the role of AI in education. By analyzing student-generated text from essays, discussion forums, and feedback, NLP tools can evaluate performance and engagement. Large language models, such as GPT and BERT, have demonstrated effectiveness in automated essay scoring and the semantic understanding of student writing, offering scalable solutions for qualitative assessment (Shermis and Burstein, 2013).

2.6. AI-Based Prediction of Student Performance Using Personality Traits

The integration of personality data into AI-based predictive models marks a significant development in educational data mining. Personality traits offer stable, theory-based variables that enhance the predictive power of traditional academic and behavioural indicators. Feature engineering efforts increasingly include the Big Five personality dimensions, which have been shown to improve model accuracy by 5–15% compared to models relying solely on academic and demo-graphic data (Okubo et al., 2017). Furthermore, AI can now infer personality traits from digital behaviours such as social media activity, learning management system interactions, and textual data, using natural language processing to extract personality indicators with reasonable accuracy (Park et al., 2015).

Methodologically, the use of ensemble methods, combining various machine learning algorithms has led to improved predictive performance when personality data is included. These methods balance the strengths of different algorithms, enhancing the robustness of predictions (Ahmad, Ismail and Aziz, 2015). Deep learning models have also incorporated personality traits into their architectures, al-lowing for the capture of complex, non-linear relationships between these traits and other student characteristics, thereby improving prediction outcomes (Hussain et al., 2018). The predictive performance of models that integrate personality data is strong, with reported accuracy rates ranging from 75% to 90% for various academic outcomes. Including personality traits typically enhances evaluation metrics such as precision, recall, and F1-scores (Karagiannis and Satratzemi, 2018). These models also show good temporal stability, maintaining accuracy across different academic terms and cohorts due to the consistency of personality traits over time (Kosinski et al., 2015).

However, the use of personality data introduces significant ethical challenges. Concerns regarding privacy, informed consent, and algorithmic bias must be addressed to ensure the responsible use of this data. Research highlights the importance of transparent algorithms, opt-in consent mechanisms, and ongoing bias audits to safeguard ethical standards in the educational applications of personality-driven AI systems (Baker and Hawn, 2022).

2.7. Research Gap: What Specific Knowledge Gap Does This Study Address?

While leadership traits and personality have been linked to academic success, there is limited research exploring the direct correlation between leadership personality traits and students' average grades.

• Most studies focus on general personality dimensions (Big Five traits) rather than leadership-specific traits such as transformational leadership behaviours, emotional intelligence competencies, and social influence capabilities.

• Limited integration of machine learning approaches exists to predict student performance based on comprehensive personality assessments. Current AI-based prediction models predominantly utilize demographic, behavioural, and basic academic variables, while underutilizing the predictive power of leadership-specific personality traits.

• There is a need for more multi-dimensional analyses, combining self-assessment results with academic data across diverse student populations. Most existing research relies on single-source data collection methods rather than comprehensive assessment frameworks that integrate multiple personality measurement approaches.

• Insufficient focus on leadership trait specificity: While general personality research in aca-demic contexts is extensive, there remains a significant gap in understanding how specific leadership characteristics (such as inspirational motivation, intellectual stimulation, individualized consideration, and idealized influence) directly correlate with academic performance outcomes.

• Lack of comprehensive machine learning modelling: Few research models systematically integrate advanced machine learning techniques with leadership personality assessments to create robust predictive frameworks for academic success.

• Limited multi-dimensional analytical approaches: Current research typically examines isolated relationships between single personality dimensions and academic outcomes, rather than exploring complex interactions and patterns through comprehensive analytical frameworks.

This study aims to address these gaps by using a comprehensive personality assessment and machine learning modelling approach, offering more profound insights into how leadership traits influence academic achievement through:

1. Direct examination of leadership-specific personality traits and their correlation with student GPA.
2. Integration of advanced machine learning algorithms to create predictive models based on leadership personality assessments.
3. Multi-dimensional analysis combining self-assessment data with academic performance across diverse student populations.
4. Development of a comprehensive framework that bridges leadership theory, personality psychology, and educational data science.

2.8. Related Work Summary

This exploration of the state-of-the-art literature reveals a rich landscape of research encompassing leadership trait theory, determinants of academic performance, relationships between personality and achievement, self-assessment methodologies, and AI-based educational prediction systems. The convergence of these fields presents significant opportunities for advancing our understanding of student success and developing more effective predictive models.

The identified research gaps highlight the need for more integrated approaches that combine leadership trait theory with advanced AI methodologies, consider cultural and contextual factors, and focus on actionable insights for educational practice. Future research should prioritize longitudinal studies that examine the dynamic relationships between personality, leadership development, and academic success, while also developing ethical frameworks for AI-based prediction systems in educational contexts. The integration of personality traits, particularly leadership-relevant characteristics, into AI-based prediction systems represents a promising avenue for enhancing student success prediction and intervention. However, realizing this potential requires addressing the methodological, ethical, and practical challenges identified in this review while maintaining focus on the ultimate goal of improving educational outcomes for all students.

3.0. Materials and Methods

For the research design, to achieve the objective of this study, we employed a quantitative approach, which involved the use of techniques such as Exploratory Data Analysis, Correlation Analysis and Machine Learning (ML) techniques to determine academic success based on personality variables quantified through several personality tests. 129 samples of MSc student assessments of the leadership module collected in the Environ-mental Engineering Department. The sample is geographically diverse, comprising students from 17 countries. The male/female ratio is 1.3/1. Twenty-three (23) personal traits were investigated through self-analytical tools, such as Personality Insight, Workplace Culture, Motivation at Work, Management Skills, and Emotion Control. Personality Inside test checks for extraversion, agreeableness, emotional stability, conscientiousness, and openness to experience. Workplace Culture tests for immediate, entrepreneurial/creative, family, achievement, mission and bureaucratic culture. Motivation at Work focuses on development, purpose, control, recognition, status, failure aversion, reward, Achievement, stability, and interaction. Academic performance metrics (aver-aging students' performance) were collected from academic reports. The feature description used in the study is provided in Table 1.

Table 1. Variable/feature names and their characteristics.

Variable Names	Characteristics
Motivation at Work	
Development	The need for personal and professional growth, learning, and improvement.
Purpose	The desire to do meaningful work that aligns with personal values or contributes to a greater good.
Control	The need for autonomy and the ability to influence decisions that affect one's work.
Recognition	The need to be acknowledged or praised for contributions and achievements.
Status	The motivation to attain prestige, rank, or reputation in an organization or field.
Failure Aversion	Motivation driven by a fear of failure often leads to over-preparation or risk avoidance.
Reward	Motivation is based on tangible benefits, such as salary, bonuses, perks, or other external incentives.
Achievement	A drive to excel, master tasks, and accomplish challenging goals.
Stability	The desire for predictability, job security, and consistency.
Interaction	The motivation that comes from social relationships and teamwork.
Personality Inside	
Extraversion	The extent to which someone is outgoing, energetic, and sociable.
Agreeableness	The degree of a person's compassion, cooperativeness, and consideration toward others.
Emotional Stability	The ability to remain calm, resilient, and free from persistent negative emotions.
Conscientiousness	Reflects how organized, responsible, and goal-oriented a person is.
Openness to Experience	Describes how creative, curious, and open-minded someone is.
Workplace Culture	

Immediate Culture	A fast-paced, results-driven environment focused on quick decisions and rapid responses.
Entrepreneurial / Creative Culture	A culture that encourages innovation, experimentation, and risk-taking.
Family Culture	A collaborative and supportive environment that emphasizes loyalty, trust, and a sense of belonging.
Achievement Culture	Focused on high performance, competition, and goal achievement.
Bureaucratic Culture	A structured, rules-based environment with clear procedures and hierarchy.
Mission Culture	Driven by a strong sense of purpose, values, and organizational goals.
Management Skills	An individual's ability to lead, organise, make decisions, communicate effectively, and manage people and resources effectively.
Emotional Control	An individual's ability to recognise, manage, and respond to emotional experiences, especially under stress or pressure, is often linked to Emotional Intelligence (EQ).

3.1. Dataset preparation procedures

After data retrieval, the results from the self-assessment test were scaled between 1 (low) and 5 (high) and visualised in histograms. Management skills results were ranked between 0 and 130, and the emotional control tests were ranked between 0 and 100. Those results, together with the Leadership in the Digital Age mark, were matched to the average grades of each student. The final dataset comprised 129 participants and included 25 variables. Written consent regarding the use of data was obtained from the students. The dataset preparation procedure adhered to ethical considerations and data privacy protection policies. Personal information, such as student names and student IDs, was anonymized.

3.2. Data analysis and modelling methods

The tests were done on Python. For the results' visualization were used the following libraries: seaborn and matplotlib. Feature selection was based on the values of their correlation coefficients. Sixteen features with the coefficients 0.05 or above were selected for machine learning modelling. The average grades were separated into three categories: "Fail" with a score below 50 out of 100, "Pass" with a score between 50 and 70, and "Excellent" with a score above 70. Seven ML algorithms, such as Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbours (KNN), Decision Tree (DT), Gradient Boosting (GB), Random Forest (RF), XGBoost (XGB) and LightGBM (LGBM), were tested in the study. Hyperparameter search grid (GridSearchCV) was used to optimize the performance of each based model. The procedure was accompanied with 5-fold cross-validation. The final best grids used for models' training are summarized in Table 2.

Table 2. Optimal hyperparameters are used for modelling.

ML Algorithm	Hyperparameters
SVM	{'C': 10, 'kernel': 'rbf'}
LG	{'C': 10, 'penalty': 'l2', 'solver': 'lbfgs'}
KNN	{'n_neighbors': 5, 'weights': 'distance'}
DT	{'max_depth': None, 'min_samples_split': 5}
GB	{'max_depth': 7, 'n_estimators': 200}
RF	{'max_depth': 10, 'min_samples_split': 2, 'n_estimators': 200}
XGB	{'max_depth': None, 'min_samples_split': 2, 'n_estimators': 200}
LGBM	{'max_depth': None, 'min_samples_split': 2, 'n_estimators': 200}

4.0. Results

4.1. Descriptive statistics and correlation analysis

The dataset was checked for missing, duplicate values, and outliers. A clean dataset was used for the correlation analysis of personal leadership traits and academic performance. Figure 1 reflects scores of 23 normalized personal traits after data standardization using a standard scaler.

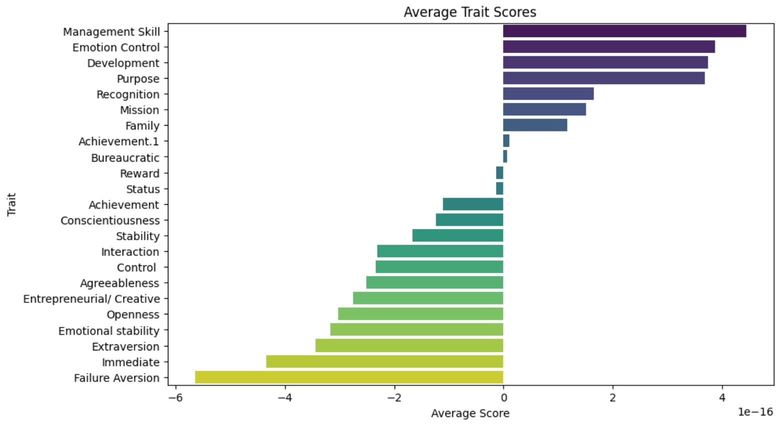

Figure 1. Personal traits scaling after standardization procedure.

This horizontal bar chart shows average trait scores across various personality and behavioural characteristics. Management Skill, Emotion Control, and Development rank highest with positive scores around 4-5, while middle-tier traits like Purpose and Recognition show moderate positive scores. Lower-scoring traits include Agreeableness and Emotional Stability, with Failure Aversion showing the most negative score at approximately -4. The chart uses a colour gradient from purple (highest) to yellow (lowest) to distinguish performance levels. Figure 2 illustrates the correlation between individual leadership traits and leadership performance ratings.

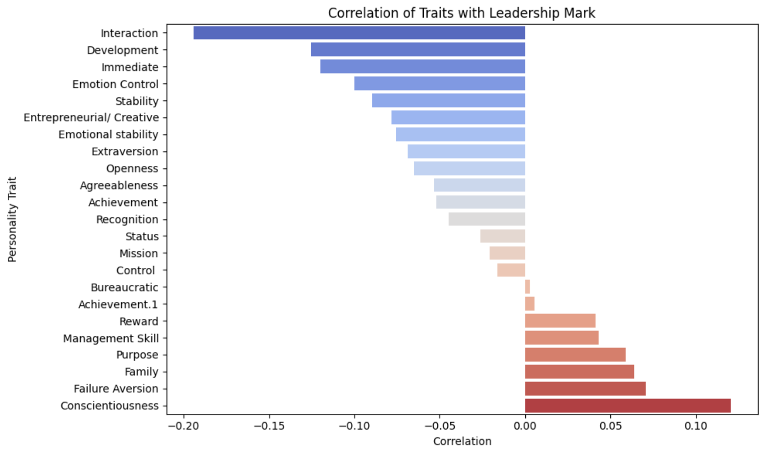

Figure 2. Correlation dependencies between personal traits and leadership mark.

This horizontal bar chart displays the correlation of various personality traits with Leadership Mark. Traits are ordered from highest positive correlation (Interaction, Development, Immediate) at the top in blue, through neutral correlations in the middle, to negative correlations at the bottom in red. Conscientiousness and Failure Aversion show the strongest negative correlations with leadership, while traits like Management Skill, Purpose, and Family also correlate negatively. The correlation values range from approximately -0.20 to +0.15.

The highest correlation coefficients (below -0.1 and above 0.1) were identified between leadership marks and variables such as "interaction", "development", "immediacy", "emotional control", and "conscientiousness". It is important to note that among the mentioned features, only "conscientiousness" shows a positive correlation with the mark. Similarly, the correlation between the average grade and personal traits was checked. The results of it are reflected in Figure 3**.**

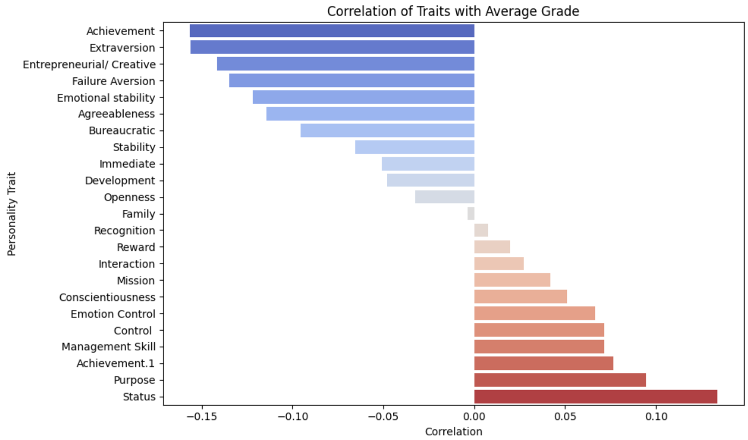

Figure 3. Correlation dependencies between personal traits and average grade.

This horizontal bar chart shows the correlation of personality traits with Average Grade. Achievement, Extraversion, and Entrepreneurial/Creative traits show the strongest positive correlations with grades (around 0.10-0.15) in blue. Most traits in the middle show weak or near-zero correlations. At the bottom, traits like Purpose, Achievement-1, Management Skill, and Status display negative correlations with academic performance in red, with Status showing the strongest negative correlation at approximately -0.10.

The strongest correlation coefficients were identified between (below -0.1 and above 0.1) "achievement", "extraversion", "entrepreneurial/creative", "failure aversion", "emotional stability", "agreeableness", and "status". The first six features in the row have a negative correlation with performance, and the only "status" feature demonstrates a positive effect on performance. The correlation coefficients for each feature are given in Table 3.

Table 3. Feature correlation coefficients.

Variable Name	Pearson Correlation Coefficient
Averages	1.000000
Leadership Mark	0.265457
Status	0.133898
Purpose	0.094405
Achievement.1	0.076527
Management Skills	0.071644
Control	0.071583
Emotion Control	0.066421
Conscientiousness	0.051116
Mission	0.041769
Integration	0.027280
Reward	0.019607
Recognition	0.007730
Family	-0.003501
Openness	-0.032390
Development	-0.048041
Immediate	-0.050725
Stability	-0.065290
Bureaucratic	-0.095548
Agreeableness	-0.114376
Emotional Stability	-0.122030
Failure Aversion	-0.134780
Entrepreneurial/Creative	-0.141348
Extraversion	-0.156131
Achievement	-0.156720

4.2. Feature selection and dataset balancing

The feature selection procedure was based on feature filtering, using a correlation coefficient threshold below -0.05 and above 0.05. 18 features were selected for the modelling. Before data modelling, a new column called "Performance" was created. The "Performance" column included three groups (classes) of the results based on averaging values: "Fail", "Pass" and "Excellent". The "Averages" variable was dropped. After that, the classes of the dataset were balanced using the Synthetic Minority Over-sampling Technique (SMOTE). The original data has a shape of 129x18, while the resampled data has a shape of 276x18. Class distribution before SMOTE: class 0 – 12 samples, class 1 – 25 samples, class 2 – 92 samples. Class distribution after SMOTE: class 0 – 92 samples, class 1 – 92 samples, class 2 – 92 samples.

4.3. Data modelling

The data was separated into training, validation, and testing parts in a proportion of 80:10:10. Table 4 summarises the machine learning (ML) performance metrics for the testing dataset of SVM, LR, KNN, DT, GB, RF, XGB, and LGBM.

Table 1. ML models classification performance for three categories of grades.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.875000	0.893856	0.875000	0.875662
XGBoost	0.875000	0.875642	0.875000	0.872126
Gradient Boosting	0.857143	0.882311	0.857143	0.856874
LightGBM	0.857143	0.866562	0.857143	0.855138
Support Vector Machine	0.839286	0.868452	0.839286	0.832172
Logistic Regression	0.821429	0.821429	0.821429	0.821429
Decision Tree	0.767857	0.777381	0.767857	0.770364
K- Nearest Neighbours	0.696429	0.672958	0.696429	0.656341

Overall, the best model for the current dataset is Random Forest, which demonstrates 87.50% accuracy, 89.38% precision, 87.50% recall, and 87.56% F1-score. Very close to RF performance, XGBoost achieves an accuracy of 87.50%, precision of 87.56%, recall of 87.50%, and F1-score of 87.21 (see Figure 4). The less effective models in classification tasks are DT and KNN.

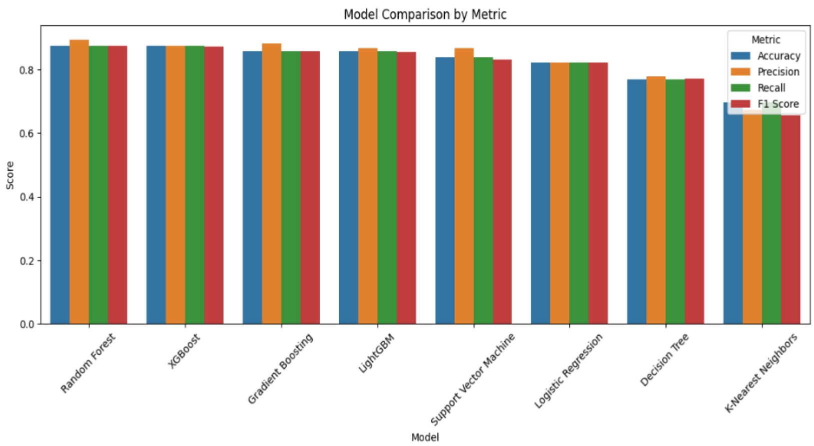

Figure 4. Performance metrics in comparison.

This grouped bar chart compares the performance of nine machine learning models across four evaluation metrics: Accuracy, Precision, Recall, and F1 Score. Most models achieve scores between 0.7-0.9, with Random Forest, XGBoost, and Gradient Boosting showing the highest performance across all metrics. Decision Tree and K-Nearest Neighbour display the lowest scores, around 0.6-0.7. The metrics are represented by different coloured bars (blue, orange, green, and red) for each model, showing relatively consistent performance patterns across the different evaluation measures.

An additional test for dataset modelling was conducted without the "Leadership Mark" variable. Table 5 demonstrates the results of modelling based on 17 personal traits features for the testing set.

Table 2. Model's performance for test for 17 leadership traits and 3 categories of grades.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.857143	0.879104	0.857143	0.857560
XGBoost	0.821429	0.828508	0.821429	0.823872
LightGBM	0.803571	0.818027	0.803571	0.805883
Support Vector Machine	0.803571	0.828125	0.803571	0.782710
Gradient Boosting	0.767857	0.773101	0.767857	0.770135
Decision Tree	0.750000	0.751284	0.750000	0.748035
Logistic Regression	0.732143	0.719345	0.732143	0.719239
K- Nearest Neighbours	0.714286	0.695727	0.714286	0.680463

The best model, similar to the first test, is RF, with an accuracy of 85.71%, a precision of 87.91%, a recall of 85.71%, and an F1-score of 85.75%. The XGB model remains in second place with an accuracy of 82.14%, precision of 82.85%, recall of 82.14%, and F1-score of 82.38%. The worst-performing models are LR and KNN.

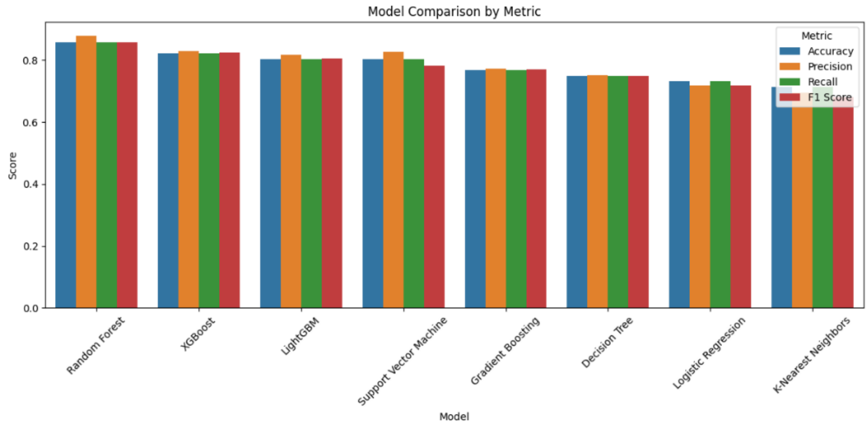

Figure 5 illustrates performance metrics in comparison.

Figure 5. Performance metrics in comparison for 17 features and 3 classes. This grouped bar chart compares the performance of eight machine learning models across four evaluation metrics: Accuracy, Precision, Recall, and F1 Score. Random Forest and XGBoost achieve the highest performance with scores around 0.85-0.9 across all metrics. Most models perform in the 0.7-0.8 range, while Decision Tree, Logistic Regression, and K-Nearest Neighbor show the lowest scores around 0.65-0.75. The four metrics are represented by different colored bars (blue, orange, green, and red), with most models showing consistent performance across all evaluation measures.

All the performance results were validated using a 5 cross-validation procedure.

4.4. Feature Importance

The feature importance for the RF, XGB, LGBM, GB, DT, and LR models is shown in Figure 6.

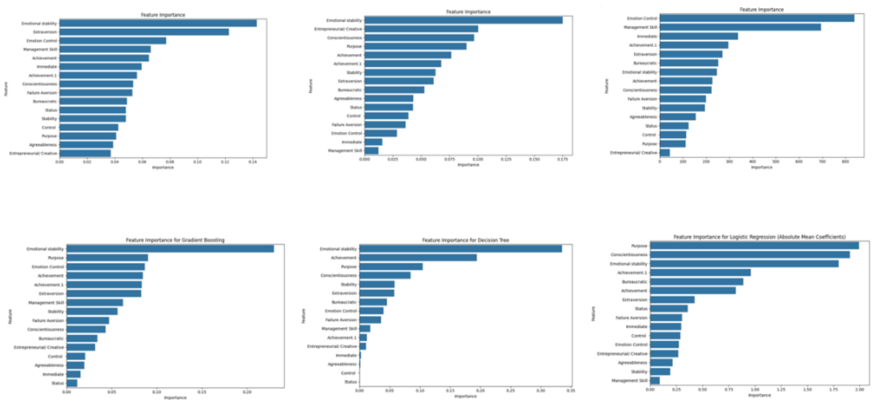

Figure 6. Feature importance for machine learning modelling.

This figure displays six horizontal bar charts showing feature importance rankings across different machine learning models or analyses. Each subplot contains multiple features ranked by their importance scores, with longer blue bars indicating higher importance. The charts appear to compare how different features contribute to model performance, with each panel showing a distinct ranking pattern. The features are listed on the y-axis while important values are shown on the x-axis, allowing for easy comparison of which variables are most influential in each respective model or analysis.

For SVM and KNN models, feature importance was not directly available. Table 6 highlights the five most essential features in the models.

Table 3. Feature importance for modelling.

RF	XGBoost	LGBM	GB	DT	LR
Emotional Stability	Emotional Stability	Emotional Control	Emotional Stability	Emotional Stability	Purpose
Extraversion	Entrepreneurial/ Creative	Management Skill	Purpose	Achievement	Conscientiousness
Emotion Control	Conscientiousness	Immediate	Emotion Control	Purpose	Emotional Stability
Management Skill	Purpose	Achievement.1 (Achievement Culture)	Achievement	Conscientiousness	Achievement.1 (Achievement Culture)
Achievement	Achievement	Extraversion	Achievement.1 (Achievement Culture)	Stability	Bureaucratic

Feature importance indicates that emotional stability is the most valuable feature, as it appears in all 6 models in the table. The second most frequent features are achievement, achievement culture, and purpose (3 out of 6), followed by emotional control and conscientiousness (2 out of 6).

4.0. Discussion and Conclusions

The study demonstrates the high predictive power of AI technologies for average students' grades based on their leadership traits. The study presents an additional opportunity to identify students' strengths and weaknesses at an early stage and choose the correct strategies for personalized learning.

Compared to the existing literature, which has primarily focused on general personality dimensions (e.g., Big Five traits) and their impact on students' performance [44, 45], this study ex-amines the influence of leadership traits. Most of the existing research relies on single-source data. In contrast, the current study combined five leadership personality tests. The state-of-the-art literature involving AI technology reflects behavioural patterns [37, 38, 40] or relies on text analysis for grade prediction [43]. A minimal number of studies describe the predictions of student grades based on personality types, and none of them were focused on the performance predictive modelling using leadership traits.

Eighteen for the first model and seventeen personality traits for the second model were used in the current research. The highest training and testing performances were achieved with RF, with an accuracy of 87.50% for the model that included leadership mark for the model building, and an accuracy of 85.71% for the model without it. The accuracy rates for academic outcome prediction using heterogeneous data vary between 75% and 90% in the state-of-the-art literature. From this perspective, our predictions are highly accurate and have the potential to be further improved.

The research helped provide answers to the research questions.
How do leadership personality traits correlate with students' academic performance?
The students' academic performance shows a direct linear correlation between leadership personality traits and average grades.
Which leadership personality traits appear most influential in predicting higher student grades?
The strongest correlation coefficients were identified between the average grade and traits such as achievement, extraversion, entrepreneurial/creative culture, failure aversion, emotional stability, agreeableness, and motivation to attain a status. At the same time, the most influential features in machine learning modelling were leadership traits, including a drive to accomplish challenging goals (achievement), a focus on high performance, competition, and goal achievement (achievement culture), a desire to do

meaningful work (purpose), emotional control, and conscientious-ness. Most of all, it is essential to note that some of the features have a positive impact on the average performance, and some have an adverse effect. We expect that feature importance can vary between university departments and reflect the differences between students in STEM and humanities subjects. Further research can give a deeper insight into personality traits and their effect on student achievements.

Can a machine learning model accurately predict students' grade categories based on leadership personality traits?

Three categories of performance, “Pass”, “Excellent”, and “Fail”, were identified using a Random Forest model with an accuracy of 85.71%, precision of 87.91%, recall of 85.71% and F1-score of 85.75%.

How can universities use insights from this study to develop personalized academic strategies tailored to students' strengths and personality profiles?

Comprehensive leadership trait development strategies can be incorporated into university pro-grams. For example, building an intrapreneurial culture and developing a teaching strategy on how to accomplish challenging goals and develop emotional stability during the project development cycle. At the same time, students who have insufficient or undeveloped leadership traits can create a personal development plan to address weaknesses and strengthen traits that contribute to higher marks. Students can be offered stress management programs. Additionally, students with a high probability of failing should receive extra academic and psychological support tailored to their personality types.

The suggested strategies can be based on student trait profiles. For example, an introverted student with a high level of conscientiousness will be assigned to an independent research or project-based task. A student with a high level of openness, who prefers a low-structured environment, will be recommended interdisciplinary electives and creative assessments. A highly extraverted person with low self-discipline will be allocated to collaborative study groups with structured timeframes. Finally, a student with high neuroticism and low resilience will be paired with a mentor and included in the stress management workshops.

The course of study will be offered, structured and dynamically adjusted using properties of ML models and students' personality features. Course recommendations will be tuned according to the predicted difficulty. A personality profile might help with selecting the assignment types (e.g., projects vs. timed tests) and with instruction formats (e.g., visual vs. audio vs. textual content). Despite promising findings of the current research, data collection was limited by the EE department (e.g., business management, humanities, and other departments were not included in the study at this stage). The sample size could also be extended.

Future research directions could utilize additional features obtained from career development tests (e.g., personal resilience, learning style, or sound decision-making) and, together with leadership traits, explore their effects on students' performance. Modelling of student performance, including that of other university departments and pro-grams, can provide a deeper understanding of human nature and its impact on student achievements.

While developing new programs or platforms for personalized learning (Schicchi and Taibi, 2024), the researchers and software developers must consider the complexity of the program and potential user issues with AI tools (Gunawardena, Bishop, and Aviruppola, 2024). Future research will focus on developing a teaching/learning framework that helps students personalize their learning and develop valuable skills for achieving high academic performance.

Author Contributions: Conceptualization, N.H. and R.F.; methodology, N.H.; software, N.H.; validation, N.H., D.H. and R.B.S; formal analysis, N.H.; investigation, N.H.; resources, N.H.; data curation, N.H.; writing—original draft preparation, N.H., R.B.S., and D.H; visualization, N.H.; project administration, N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data is held on private university domain.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Avolio, B.J. & Gardner, W.L. (2005) 'Authentic leadership development: Getting to the root of positive forms of leadership', *The Leadership Quarterly*, 16(3), pp. 315–338. https://doi.org/10.1016/j.leaqua.2005.03.001

Ahmad, F., Ismail, N.H. & Aziz, A.A. (2015) 'The prediction of students' academic performance using classification data mining techniques', *Applied Mathematical Sciences*, 9(129), pp. 6415–6426.

Andrade, H. & Du, Y. (2007) 'Student responses to criteria-referenced self-assessment', *Assessment & Evaluation in Higher Education*, 32(2), pp. 159–181. https://doi.org/10.1080/02602930600801928

Baker, R.S. & Hawn, A. (2022) 'Algorithmic bias in education', *International Journal of Artificial Intelligence in Education*, 32(4), pp. 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Bandura, A. (1997) *Self-efficacy: The exercise of control*. New York: Freeman.

Carlyle, T. (1993) *On heroes, hero-worship, and the heroic in history*. Vol. 1. Berkeley: University of California Press.

Chamorro-Premuzic, T. & Furnham, A. (2003) 'Personality predicts academic performance: Evidence from two longitudinal university samples', *Journal of Research in Personality*, 37(4), pp. 319–338. https://doi.org/10.1016/S0092-6566(02)00578-0

Chen, X., Zou, D., Cheng, G. & Xie, H. (2020) 'Detecting latent topics and trends in educational technologies over four decades using structural topic modelling', *Computers & Education*, 151, 103846. https://doi.org/10.1016/j.compedu.2020.103855

Chrysafiadi, K. & Virvou, M. (2013) 'Student modelling approaches: A literature review for the last decade', *Expert Systems with Applications*, 40(11), pp. 4715–4729. https://doi.org/10.1016/j.eswa.2013.02.007

Connolly, J.J., Kavanagh, E.J. & Viswesvaran, C. (2007) 'The convergent validity between self and observer ratings of personality: A meta-analytic review', *International Journal of Selection and Assessment*, 15(1), pp. 110–117. https://doi.org/10.1111/j.1468-2389.2007.00371.x

Duckworth, A.L., Peterson, C., Matthews, M.D. & Kelly, D.R. (2007) 'Grit: Perseverance and passion for long-term goals', *Journal of Personality and Social Psychology*, 92(6), pp. 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Dunning, D., Heath, C. & Suls, J.M. (2004) 'Flawed self-assessment: Implications for health, education, and the workplace', *Psychological Science in the Public Interest*, 5(3), pp. 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Dutt, A., Ismail, M.A. & Herawan, T. (2017) 'A systematic review on educational data mining', *IEEE Access*, 5, pp. 15991–16005. https://doi.org/10.1109/ACCESS.2017.2654247.

Dweck, C.S. (2006) *Mindset: The new psychology of success*. New York: Random House.

Germain, M.-L. (2008) *Traits and skills theories as the nexus between leadership and expertise: Reality or fallacy?* Academy of Human Resource Development Conference, Panama City, FL. https://files.eric.ed.gov/fulltext/ED501636.pdf

Goleman, D. (1995) *Emotional intelligence*. New York: Bantam Books.

Gunawardena, M., Bishop, P. & Aviruppola, K. (2024) 'Personalized learning: The simple, the complicated, the complex and the chaotic', *Teaching and Teacher Education*, 139, 104429. https://doi.org/10.1016/j.tate.2023.104429

Halkiopoulos, C. & Gkintoni, E. (2024) 'Leveraging AI in e-learning: Personalized learning and adaptive assessment through cognitive neuropsychology', *Electronics*, 13(18), 3762. https://doi.org/10.3390/electronics13183762

Helal, S. et al. (2018) 'Predicting academic performance by considering student heterogeneity', *Knowledge-Based Systems*, 161, pp. 134–146. https://doi.org/10.1016/j.knosys.2018.07.042

Hussain, M., Zhu, W., Zhang, W. & Abidi, S.M.R. (2018) 'Student engagement predictions in an e-learning system and their impact on student course assessment scores', *Computational Intelligence and Neuroscience*, 2018. https://doi.org/10.1155/2018/6347186

John, O.P. & Srivastava, S. (1999) 'The Big Five trait taxonomy: History, measurement, and theoretical perspectives', in *Handbook of Personality: Theory and Research*, 2nd ed., pp. 102–138.

Judge, T.A., Bono, J.E., Ilies, R. & Gerhardt, M.W. (2002) 'Personality and leadership: A qualitative and quantitative review', *Journal of Applied Psychology*, 87(4), pp. 765–780.

Judge, T.A., Colbert, A.E. & Ilies, R. (2004) 'Intelligence and leadership: A quantitative review and test of theoretical propositions', *Journal of Applied Psychology*, 89(3), pp. 542–552. https://doi.org/10.1037/0021-9010.89.3.542

Karagiannis, I. & Satratzemi, M. (2018) 'An adaptive mechanism for Moodle based on automatic detection of learning styles', *Education and Information Technologies*, 23(3), pp. 1331–1357. https://doi.org/10.1007/s10639-017-9663-5

Kirkpatrick, S.A. & Locke, E.A. (1991) 'Leadership: Do traits matter?', *Academy of Management Perspectives*, 5(2), pp. 48–60.

Kosinski, M. et al. (2015) 'Facebook as a research tool for the social sciences', *American Psychologist*, 70(6), pp. 543–556. https://doi.org/10.1037/a0039210

Luckin, R. & Cukurova, M. (2019) 'Designing educational technologies in the age of AI: A learning sciences-driven approach', *British Journal of Educational Technology*, 50(6), pp. 2824–2838.

McClelland, D.C. (1985) *Human motivation*. Glenview, IL: Scott, Foresman.

Northouse, P.G. (2021) *Leadership: Theory and practice*. 8th ed. Thousand Oaks, CA: Sage.

Okubo, F. et al. (2017) 'A neural network approach for students' performance prediction', *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pp. 598–599.

Organ, D.W. (1996) 'Leadership: The great man theory revisited', *Business Horizons*, 39(3), pp. 1–4.

Ouyang, F. & Zhang, L. (2024) 'AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review', *Educational Research Review*, 44, 100616.

Panadero, E., Jonsson, A. & Botella, J. (2017) 'Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses', *Educational Research Review*, 22, pp. 74–98.

Park, G. et al. (2015) 'Automatic personality assessment through social media language', *Journal of Personality and Social Psychology*, 108(6), pp. 934–952.

Pascarella, E.T. & Terenzini, P.T. (2005) *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

Paulhus, D.L. (2002) 'Socially desirable responding: The evolution of a construct', in *The Role of Constructs in Psychological and Educational Measurement*, pp. 49–69.

Peterson, C. & Seligman, M.E. (2004) *Character strengths and virtues: A handbook and classification*. Oxford: Oxford University Press.

Poropat, A.E. (2009) 'A meta-analysis of the five-factor model of personality and academic performance', *Psychological Bulletin*, 135(2), pp. 322–338.

Poropat, A.E. (2014) 'Other-rated personality and academic performance: Evidence and implications', *Learning and Individual Differences*, 34, pp. 24–32.

Richardson, M., Abraham, C. & Bond, R. (2012) 'Psychological correlates of university students' academic performance: A systematic review and meta-analysis', *Psychological Bulletin*, 138(2), pp. 353–387.

Roberts, B.W. & DelVecchio, W.F. (2000) 'The rank-order consistency of personality traits from childhood to old age', *Psychological Bulletin*, 126(1), pp. 3–25.

Romero, C. & Ventura, S. (2020) 'Educational data mining and learning analytics: An updated survey', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355.

Schicchi, D. & Taibi, D. (2024) 'Redefining education: A personalized AI platform for enhanced learning experiences', *Proceedings of the Second International Workshop on Artificial Intelligence Systems in Education*.

Shermis, M.D. & Burstein, J. (eds.) (2013) *Handbook of automated essay evaluation: Current applications and new directions*. New York: Routledge.

Sirin, S.R. (2005) 'Socioeconomic status and academic achievement: A meta-analytic review', *Review of Educational Research*, 75(3), pp. 417–453.

Tinto, V. (1993) *Leaving college: Rethinking the causes and cures of student attrition*. 2nd ed. Chicago: University of Chicago Press.

Taylor, S.E. & Brown, J.D. (1988) 'Illusion and well-being: A social psychological perspective on mental health', *Psychological Bulletin*, 103(2), pp. 193–210.

Trapmann, S., Hell, B., Hirn, J.O.W. & Schuler, H. (2007) 'Meta-analysis of the relationship between the Big Five and academic success at university', *Zeitschrift für Psychologie*, 215(2), pp. 132–151.

Ujkani, B., Minkovska, D. & Hinov, N. (2024) 'Course success prediction and early identification of at-risk students using explainable artificial intelligence', *Electronics*, 13(21), 4157.

Van Dierendonck, D. (2011) 'Servant leadership: A review and synthesis', *Journal of Management*, 37(4), pp. 1228–1261.

Vorobyeva, K.I. et al. (2025) 'Personalized learning through AI: Pedagogical approaches and critical insights', *Contemporary Educational Technology*, 17(2).

Zimmerman, B.J. (2002) 'Becoming a self-regulated learner: An overview', *Theory Into Practice*, 41(2), pp. 64–70.

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 17-32
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1725

Relationships between Airline Sustainability and Consumer Behaviour: An assessment of the influence of environmental awareness on the decision-making process of European airline customers

Yves Kremer
EY, Senior Consultant, Amsterdam

Eustathios Sainidis
Associate Professor, Northumbria University

Abstract
The aviation industry increasingly contributes to the global share of carbon emissions and therefore also to increasing global average temperatures. Although forecasts for growing emissions and solutions to decrease CO2 by aircraft usage are well described in the extant literature, the debate has failed to address how aviation passengers feel influenced in their choice as consumers. The study investigates possible relationships between civil aviation environmental sustainability, passenger environmental awareness and consumer behaviour. We use the European airline industry as a geographical focus of the study. Five hypotheses are developed related to (i) environmental awareness by airline passengers, (ii) the influence of sustainability on ticket booking behaviour, (iii) the influence of Sustainable Aviation Fuel and carbon offsetting on consumer behaviour and (iv) the influence of environmental awareness on airline image and customer satisfaction. To test the hypotheses, a survey method is used to gather data from airline passengers. The results show that environmental awareness is indeed increasing among airline passengers and as such has an influence on consumer behaviour. Our data indicates that the image of the sector is declining due to a perceived lack of urgency by the airlines. At the same time consumers are willing to pay higher ticket prices is airlines invest into sustainability. However, it is found that, although environmental awareness and concern are growing amongst aviation consumers, price is yet the most important factor that influences ticket booking behaviour and passenger satisfaction.

Keywords: Sustainability, Airlines, Consumer behaviour, Environmental awareness.
Wordcount: 236

1.0 Introduction

Environmental sustainability is a dominant topic around the world. In 1990, the first environmental assessment report by the International Panel on Climate Change (IPCC) was provided (IPCC, 2020), followed by the assembly of the United Nations Framework Convention on Climate Change (UNFCCC) – a convention joined by 197 nation-states – as a response to global climate change (UNFCCC, 2020). The aviation industry is also gradually more part of the climate-discussion, as expert reports (e.g. Lee, et al., 2009) found that emissions from airlines contribute to the rise in global average temperatures. The industry amounts to 2.4% of all human-produced CO2 emissions (M. Klöwer, 2021), although its contribution is calculated to increase (Rupcic, et al., 2023). This is due to the rapid growth of the sector, which noted an increase in passengers per year from 1.0 billion in 1990 to 4.2 billion in 2018 (The World Bank Group, 2020).

To decrease the emissions from aircraft travel, aviation became part of the EU Emissions Trading Scheme (EU ETS) in 2012 (European Commission, 2020), followed by various additional directives to improve the industry's climate performance (European Commission, 2020). Subsequently, the focus became the modelling of emissions' pattern (FitzGerald & Tol, 2007) or on factors that influence emission production in civil aviation (Brueckner & Abreu, 2017).

Following a review of the extant literature and professional reports we identified a gap in the role of environmental sustainability on airline consumers. Environmental awareness is growing among the public and has rapidly increased over the past fifteen years (Cohen, 2015), which resulted in a rising demand for environmental-friendly businesses (Gadenne, Kennedy, & McKeiver, 2009). Han et al. (2019) describe that due to the growing level of awareness, a pattern of environmental concern is now also present among airline customers. Although it is indicated that green products and environmental awareness in general affect consumer behaviour (McDonald & Oates, 2006), the affinity between consumer behaviour and environmental awareness is not regularly applied to the airline industry. The purpose of this research, therefore, is to investigate the influence of environmental awareness and sustainability on consumer behaviour for European airline passengers. Using a quantitative research methodology we survey-targeted civil aviation executives and airline passengers. Relationships between sustainability and consumer decision-making are drawn to add to the current body of knowledge.

2.0. Theoretical basis

In general, sustainability in civil aviation is based on the fact that aircraft usage contributes to the global production of CO2 emissions. Graver et al. (2019) indicate that flights contribute 2.4% of global CO2 emissions as result of fossil fuel to power aircraft engines. Civil aviation contributes the majority of the aviation industry's carbon footprint with 81% of the total CO2 emissions. Grimme (2008), suggests a decline in emissions is possible through efficiency improvements. However, sustainability is yet more often used by airlines as part of a strategy to gain commercial advantages rather than directly address sustainable concerns (Karaman & Akman, 2018). An explanation of this profit-led paradigm is that "sustainable governance always requires companies to integrate environmental factors with economic factors" (Salvioni, Gennari, & Bosetti, 2016). Nevertheless, Walker and Cook (2009) acknowledged that the expansion of flights in the industry is outgrowing overall efficiency improvements and which increases environmental impact (negative). In addition, progress in sustainable development in the sector faces structural delays due to technical uncertainty (Peeters, Higham, Kutzner, Cohen, & Gössling, 2016).

2.1. Developments in airline sustainability

In terms of environmental development, Grimme (2008) states that aircraft efficiency can be increased by improving aircraft and engine technology, the substitution of kerosene by biofuels and the optimization of air traffic management (ATM). European Union funded research concluded that emission-based ticket taxes on all flights to and from EU-member states can also result in lower CO2 emissions (Krenek & Schratzenstaller, 2016). From these options, various authors (e.g. Dray et al., 2010) believe that substituting kerosene with Sustainable Aviation Fuel (SAF) is the most advanced possibility to decrease carbon emissions. In 2011, the European Commission (EC) launched its Biofuel Flightpath Initiative (European Commission, 2020) and since then, the total number of flights executed on SAF amount to over 250,000 (ATAG, 2020). However, EASA (2020) expects the volume of SAF to remain limited in the short term. This is due to SAF's main disadvantage; high production costs.

Looking at SAF-production methods, the International Energy Agency (IEA) (2019) stresses that currently only HEFA-produced fuels are technologically mature and therefore the preferred principal aviation biofuel. According to IATA (2019) and biofuel producers and suppliers (e.g. BP p.l.c., 2020; SkyNRG, 2020)., HEFA-type SAF achieves a reduction in emissions by up to 80%. However, Neste Oil (2020) adds that an 80% reduction can only be achieved if the HEFA-SAF is used in neat form, which, due to regulations, is not an option in the aviation industry (European Technology and Innovation Platform, 2020).

The downside of this production method is that HEFA-SAF encounters high production costs due to aircraft engines requiring only high-quality paraffinic biofuels (Chiaramonti, Prussi, Buffi, & Tacconi, 2014). As such the corresponding high production costs result in high fuel prices for airlines (Bittner, Wallace, & Zhao, 2015). Although price information is limited and kerosene prices fluctuate, the price of HEFA-fuel tends to be up to 60% higher than the price of kerosene (International Energy Agency, 2019; Energy Post, 2019). Therefore, Kim, Lee and Ahn (2019) note that the commercial transition to alternative fuel remains somehow slow, with some authors describing the market penetration of biofuels in the EU as negligible (Prussi, O'Connell, & Lonza, 2019).

3.2. Airline consumer behaviour

Not only is sustainability a type of development in the sector, it can also be a factor that influences airline consumer behaviour. Hwang and Lyu (2020) found that green image is an important predictor of positive consumer attitudes towards the desire of choosing an environmentally friendly airline. Moreover, Abdul Rashid et al. (2014) found that Environmental-CSR has a potential impact on customers' behavioural response towards companies, with Park et al. (2015) elaborating that environmental responsibility – as well as economic and social responsibility – has a direct influence on airline customer satisfaction, which in turn shows a connection with behavioural intention. The significance of this influence is derived from a growing environmental concern among the public (Han, Yu, & Kim, 2019), causing a positive association between loyalty and social responsibility (Chen, Chang, & Lin, 2012).

However, there are more moderators of consumer behaviour in the airline industry, such as the price of a product (Dolnicar, Grabler, Grün, & Kulnig, 2011). Joseph (2020) describes that high prices are major barriers for environmental consumer behaviour and Rajaguru (2016) adds that perceived value for money and service quality are significant drivers of customer satisfaction and behavioural intention.

3.3. Consumer environmental awareness

As Han, Yu and Kim (2019) described earlier, a growing environmental concern is present among airline customers, as consumers are increasingly more aware of their impact on the environment (Yoo, Divita, & Kim, 2013). Furthermore, it is found that this type of awareness is present in multiple groups in society, regardless of factors such as level of income (Angelovska, Bilic Sotiroska, & Angelovska, 2012). In relation to growing environmental awareness, Gössling and Peeters (2009) note that to the public, the aviation industry does not seem to show a sense of seriousness regarding its contribution to global emissions. For instance, Graham and Shaw (2008) describe that European airlines with a low-cost business model do not account for the environmental damage that is caused, as this is not in line with their economic development. The level of concern among the public in relation to the climate impact of airlines in some cases even results in 'flight shame', which is derived from guilt associated with using air travel (Hasberg, 2019).

More recent research by Heikkinen (2020) elaborates that airline consumers are sceptical regarding the effectiveness of current sustainability developments, such as carbon offsetting by airlines, although he also states that people are willing to pay for flight carbon offsets, such as the usage of SAF or fuel-efficient aircraft. In addition, Kumar, Garg and Makkar (2012) found that people are willing to pay up to 20 per cent higher prices for green products in relation to non-green equivalents. However, regardless of consumer payments, Bhate (2001) states that environmental deterioration can only be halted if the environmental awareness of consumers is supplemented by governmental policies. According to Yang and Chen (2018), it is correct that the willingness of environmental investment by companies is altered by both environmental awareness and carbon taxes.

3.0. Methodology

The research philosophy is based on positivism, following best practice by other studies in the discipline (Burrell & Morgan, 2016). In addition, the paper uses a deductive approach towards the research, which develops one or more hypotheses based on existing theory (Nelson, et al., 2004; Wilson, 2014). The next paragraphs discuss a corresponding research strategy and data analysis method.

3.1. Research Strategy

To achieve the research objective, the research strategy is designed as single-method quantitative research, which incorporates a single research method to investigate the research subjects (Lewis-beck, Bryman, & Futing Liao, 2004). The quantitative approach is only a good fit with an interpretivist philosophy if the collected data is based on opinions (Saunders, Lewis, & Thornhill, 2019). Newman (2000) adds that a deductive approach, in general, fits a quantitative research strategy, as also stated by Wilson (2014). Therefore, a survey research strategy is the principal data collection method in this quantitative research. This strategy is used, as Saunders et al. (2019) state that the collected data from respondents can indicate relationships between variables. Singleton and Straits (2009) state that surveys are often used in describing human behaviour, with Ponto (2015) adding that surveys help to identify consumer patterns. The survey research strategy is executed through the distribution of questionnaires.

3.2. Questionnaire

The questionnaire is indicated as a "self-complemented internet questionnaire" (Saunders, Lewis, & Thornhill, 2019), as it is only distributed online. Furthermore, the foundation of the questionnaire is based on the research objectives and the literature review. Only closed-type questions are used, as these are indicated by Saunders et al. (2019) to decrease the completion time for respondents. In addition, the questions are divided over five topics: consumer environmental awareness, consumer ticket booking, biofuel, consumer carbon offsetting and customer experience & product quality. Within these subjects, both multiple-choice questions and Likert-scale questions are used, which uses a level of agreement scale in five points: (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; (5) Strongly agree" (Watson, 2010).

A data requirements table (Appendix A) is created to increase quality assurance and ensure data integrity (Smadi, et al., 2015; Saunders, Lewis, & Thornhill, 2019), which shows 30 questions, including five general questions to structure respondents. At the start of the survey, the subject of research and privacy notices from Northumbria University are stated. Furthermore, the form is perfected by performing pilot tests, which resulted in adjustments regarding the clarity of the questionnaire structure and the descriptions of terms as carbon offsetting and Sustainable Aviation Fuel.

The data sample used included air travel consumers within the European aviation industry. For this research, a person is considered a European air travel consumer if that person uses air travel one or more times per year within the EU. Consumers are structured per air travel frequency, purpose of air travel, airline preference, age and gender. In total, 200 respondents were targeted and 150 responses were gathered using a non-probability sampling technique. This includes snowball sampling (Wilson, 2014) and the use of social media platforms & messaging application WhatsApp to distribute the questionnaires.

According to Saunders et al. (2019), the collected data is only representative of the target population if the percentage of non-biased responses is over 80 per cent. As all questions are designed as mandatory in the questionnaire, it is not possible to provide a non-response and a complete response of over 80 per cent is achieved.

3.3. Data Analysis

Blaikie (2003) states that the analysis of sample data is an important process in achieving research objectives. This paper uses descriptive statistics to analyse data, which are "brief descriptive coefficients that summarize a given data set," (Kenton, 2019). Earlier, it was established that a deductive approach creates hypotheses based on current theory (Wilson, 2014). As per Glen (2020), the results of a questionnaire can be used to assess the validity of hypotheses by applying statistical hypothesis testing. It is added that this form of testing assesses relationships within sample data (Saunders, Lewis, & Thornhill, 2019) and it also is a good method for quantifying questions and answers (Brownlee, 2020). In this research, statistical hypothesis tests are performed in Microsoft Excel.

The first step in hypothesis testing is by stating a null hypothesis (H_0), which is the opposite of the research hypothesis (Johnson, 1999). Others (Helmenstine, 2020; Hayes & Westfall, 2020) add that a null hypothesis states that there is no relationship between two measured variables. Opposed to H_0 is an alternative hypothesis (H_1), describing that a relationship does exist between variables and fits better with the research hypothesis (Blaikie, 2003). As a null hypothesis can only be disproved (Johnson, 1999), null hypotheses are either rejected or failed to reject by a hypothesis test (Wilson, 2014).

The statistical tool that is used in this paper is the one-sample t-test, which "tests the null hypothesis that the mean (μ) of the population sampled is a specific value against an alternative value" (Gastwirth & Rubin, 1971). In other words, the test "determines whether the sample mean is statistically different from a known or hypothesized population mean" (Kent State University, 2020). The t-value, which is the outcome of a t-test, is compared to the critical value of the sample data to reject or fail to reject a null hypothesis. Although this is debated widely in the literature, one-sample t-tests have been justified as appropriate and sufficient for correlation analysis and to reassure evidence for the null hypotheses of no differences between two measures (Phylactou, et al, 2025; Francis and Jakicic, 2022). The outcome of a t-test is the t-statistic, which includes the following two formulae (NCSS, 2020):

$$t = \frac{\bar{x} - \mu}{\sigma \bar{x}}$$

Equation 1: Formula to calculate t-statistic value (NCSS, 2020)

$$\sigma \bar{x} = \frac{\sigma}{\sqrt{n}}$$

Equation 2: Formula to calculate standard error of mean (NCSS, 2020)

Table 1 below shows the variables that are used in the calculation of the t-statistic.

Variable	Symbol
Sample mean	$\bar{x}$
Proposed constant for the population mean	μ
Sample size	n
Sample standard deviation	σ
Standard error of mean	$\sigma\bar{x}$
Degrees of Freedom	df
Significance level	α

Table 1: Variables and symbols used in t-test related formulae

According to Brownlee (2020), a null hypothesis is rejected if the t-statistic is greater than the critical value of the sample data and a test fails to reject H_0 if t-statistic is equal or smaller than the critical value. The critical value of t is depending on the degrees of freedom (n-1) and the level of significance (Glen, 2020). Brownlee (2020) also indicates that a hypothesis can be rejected if the p-value is lower than the level of significance. The p-value "determines the probability of observing a more extreme test statistic in the direction of the alternative hypothesis" (The Pennsylvania State University, 2020). In addition, it is indicated that "the smaller the p-value, the stronger the evidence that the null hypothesis should be rejected" (Glen, 2020). Following the literature, two formulae are used to reject H_0:

- Reject H0 if t-statistic > critical value
- Reject H0 if p-value < α

To perform a statistical test, the sample data from the surveys are transformed. As per Frost (2020), the numerical coefficient of Likert scale data ascends from 1 (strongly disagree) to 2 (disagree), 3 (neutral), 4 (agree) and 5 (strongly agree). Furthermore, the proposed constant for the population mean and the level of significance maintain fixed values, as well as the null hypothesis and alternative hypothesis of all research hypotheses. A significance level of 0.05 is a desired value for academic research (Yale University, 2020; Lund Research, 2020), which is also used in this dissertation. Furthermore, the proposed constant for the population mean is set at 3, which is the mean of the five possible numerical Likert-scale answers in the questionnaire. Therefore, the null hypothesis and alternative for all research hypotheses are:

Hypotheses	
Null hypothesis (H0)	$\mu \leq 3.00$
Alternative hypothesis (H1)	$\mu > 3.00$

Table 2: The null hypothesis and alternative hypothesis for hypotheses testing

3.4. Hypotheses

Five research hypotheses are derived from the literature review and respectively correspond to the five questionnaire subjects earlier in the methodology. As indicated by Johnson (1999), a null hypothesis is the opposite of a research hypothesis and "states that there is no relationship between two measured variables" (e.g. Helmenstine, 2020). As a result, the five null hypotheses to be tested with corresponding alternative hypotheses are shown in Table 3.

Hypotheses 1-5	
$H1_0$	European airline consumers think that there is no relationship between increasing airline emissions and rising global average temperatures
$H2_0$	There is no relationship between environmental ticket-booking factors, such as flight taxes and ecolabels, and the behavioural intention of consumers to book a flight ticket
$H3_0$	There is no relationship between the behavioural intention of consumers to pay a higher price for a flight ticket and the use of biofuel instead of kerosene by airlines
$H4_0$	There is no relationship between the behavioural intention of consumers to pay a higher price for a flight ticket and the applicant of carbon offsetting options on flight tickets
$H5_0$	There is no relationship between growing environmental awareness among European airline consumers and the image & quality perception of airlines

Table 3: The null hypotheses used for hypothesis testing

3.5. Conceptual Framework

An expectation of the research findings is derived from the literature review and is expressed through a conceptual framework (Figure 1), which is used "to explain the relationships between the main variables in the study" (Adom, Hussein, & Agyem, 2018). The variables describe the relationship between 'Environmental Sustainability Factors' and 'Airline Consumer Behavioural Intention'. It is expected that the influence of sustainability factors on consumer behaviour is positively moderated by consumer environmental awareness and is negatively moderated by non-environmental factors such as ticket price. Furthermore, it is expected that the research finds that customer satisfaction and airline image are both mediating variables with a positive effect on the environmental sustainability-consumer behaviour relationship.

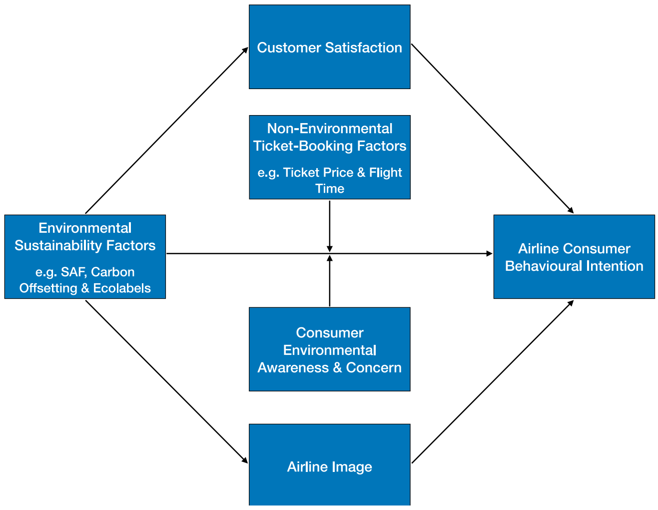

Figure 1: Conceptual framework of research variables from theory & literature review

4.0. Results

The results from research are elaborated upon in two sections; the survey results and the results from hypothesis testing.

4.1. Survey Results

The methodology describes five questionnaire subjects, which respectively correspond to the five research hypotheses. The survey results are first provided according to these subjects, which are:

1) Environmental awareness
2) Environmental sustainability during ticket booking
3) Sustainable Aviation Fuel
4) Carbon offsetting
5) Influence of sustainability on customer satisfaction and airline image

Environmental Awareness

Q1: Statement: Carbon emissions contribute to rising global average temperatures.

Q2: Statement: Environmental awareness is growing among the general public.

Q3: Statement: Airlines contribute to the increase of global carbon emissions and therefore the rising global average temperatures.

Q4: Statement: Airlines understand the seriousness of rising global average temperatures and carbon emissions

Q5: Statement: The development of greener aviation is progressing fast enough

Answers	Question 1	Question 2	Question 3	Question 4	Question 5
Strongly Agree (%)	39%	23%	26%	7%	3%
Agree (%)	44%	59%	53%	40%	10%
Neutral/Not Sure (%)	13%	13%	14%	33%	30%
Disagree (%)	3%	5%	6%	18%	43%
Strongly Disagree (%)	1%	1%	1%	2%	15%
Total (%)	100%	100%	100%	100%	100%

Table 4: Answers to survey questions 1-5 (%)

Environmental Sustainability during Ticket Booking

Q6: What factors are important to you when booking a ticket? (multiple answers possible)

Q7: If an airline is known for its contribution to environmental sustainability, would that affect your decision to book a ticket with them?

Q8: If tickets would have product ecolabels, would this influence your choice of ticket?

Q9: Statement: Implementing ticket taxes would decrease airline carbon emissions.

Q10: Statement: Flight tax should be added to flight tickets.

Answers	Question 6 (nr.)	Question 6 (%)
Ticket price	138	92%
Airline preference	65	43%
Airline image	35	23%
Service quality	71	47%
Previous experiences	91	61%
Amenities included in your ticket	22	15%
Sustainability of airlines	22	15%
Flight time/layover or direct flight	113	75%

Table 5: Answers to survey questions 6

Answers	Question 7	Question 8	Questions 9	Question 10
Strongly Disagree (%)	n/a	n/a	10%	10%
Disagree / No (%)	23%	27%	23%	15%
Neutral / Not sure (%)	8%	7%	31%	23%
Agree / Yes, but not significantly (%)	54%	48%	27%	33%
Strongly Agree / Yes, significantly (%)	15%	18%	9%	19%
Total (nr.)	100%	100%	100%	100%

Table 6: Answers to survey questions 7-10 (%)

Sustainable Aviation Fuel

Q11: Did you know that biofuel emits up to 80 per cent less CO2 compared to the current jet kerosene?

Q12: Did you know that airlines are already using biofuel on some flights?

Q13: Statement: Biofuel fuel should be used more often by airlines.

Q14: Are you willing to pay a higher ticket price if a flight uses biofuel instead of kerosene?

Q15: Would you choose an airline because it uses more biofuel than other airlines do?

Answers	Question 11	Question 12	Question 13	Question 14	Question 15
No, I do not believe it / Strongly Disagree	6%	n/a	3%	n/a	n/a
No, I did not know before / No / Disagree	63%	49%	1%	19%	29%
Neutral / Not sure	6%	8%	15%	9%	25%
Yes (Q1,2,5) / Yes, but up to 10% higher / Agree	n/a	43%	43%	49%	47%
Strongly Agree / Yes (Q4)	25%	n/a	39%	23%	n/a
Total	100%	100%	100%	100%	100%

Table 7: Answers to survey questions 11-15 (%)

Carbon Offsetting

Q16: Do you use carbon offsetting options when booking a ticket?

Q17: Statement: Letting consumers pay for carbon offsetting is an effective method in decreasing airline emissions.

Q18: Do you think paying for carbon offsetting should be voluntary or mandatory for customers?

Q19: Are you willing to pay a higher ticket price if mandatory carbon offsetting is included?

Q20: Would you be more willing to pay for carbon offsetting if you know how your payment is used by airlines?

Answers	Question 16	Question 17	Question 18	Question 19	Question 20
Strongly Disagree (%)	n/a	11%	n/a	n/a	n/a
Disagree / No / Voluntary (%)	61%	24%	38%	26%	10%
Neutral / Not sure (%)	11%	37%	15%	5%	21%
Agree / Yes, Sometimes / Mandatory / Yes, but up to 10% higher / Yes (Q5) (%)	19%	22%	47%	49%	69%
Strongly Agree / Yes, Always / Yes (Q4) (%)	9%	6%	n/a	20%	n/a
Total (%)	100%	100%	100%	100%	100%

Table 8: Answers to survey questions 16-20 (%)

Influence of Sustainability on customer satisfaction and airline image

Q21: Statement: Environmental Sustainability by airlines has an influence on my customer satisfaction.
Q22: Statement: The price of a flight ticket has an influence on my customer satisfaction.
Q23: Statement: I believe airlines that claim to run a CO2 neutral operation.
Q24: Statement: The image of airlines is decreasing because of growing environmental awareness.
Q25: Statement: I experience a feeling of guilt in taking a flight.

Answers	Question 21	Question 22	Question 23	Question 24	Question 25
Strongly Agree (%)	10%	28%	1%	11%	6%
Agree (%)	39%	62%	18%	48%	20%
Neutral/Not Sure (%)	30%	9%	30%	20%	18%
Disagree (%)	17%	1%	40%	20%	38%
Strongly Disagree (%)	5%	1%	11%	1%	18%
Total (%)	100%	100%	100%	100%	100%

Table 9: Answers to survey questions 21-25 (%)

4.2. Hypothesis Testing Results

The research methodology describes that the indicated hypotheses are tested by performing statistical t-tests. The data used to perform these tests is derived from the survey results and are divided per questionnaire subject/hypothesis. However, not every question within a subject is applicable to the corresponding hypothesis. Therefore, the count (n) used to calculate the t-statistic varies per hypothesis.

As seen in the hypotheses testing dataset in Table 10, Hypotheses 1 & 4 use the answers of all questions, accumulating of a total count (n) of 750 each. The second and fifth hypothesis both use a count of 600, as respondents were able to provide multiple answers to question 6 and the topic of question 22 does not comply with the topic of the last hypothesis. The lowest count is used for Hypothesis 3, because the first two questions only tested the knowledge of the respondents and answers are therefore left out of the dataset.

Hypothesis	Coefficient	Data measurement detail	Answers (Nr.)
Hypothesis 1	1	Strongly Disagree	28
	2	Disagree	112
	3	Neutral	153
	4	Agree	310
	5	Strongly Agree	147
	Count (n)	**Total**	**750**
Hypothesis 2	1	Strongly Disagree	42
	2	Disagree / No	131
	3	Neutral / Not sure	104
	4	Agree / Yes, but not significantly	244
	5	Strongly Agree / Yes, significantly	79
	Count (n)	**Total**	**600**
Hypothesis 3	1	Strongly Disagree	4
	2	Disagree / No	72
	3	Neutral / Not sure	74
	4	Agree / Yes, but up to 10% higher / Yes (Q5)	207
	5	Strongly Agree / Yes (Q4)	93
	Count (n)	**Total**	**450**
Hypothesis 4	1	Strongly Disagree	16
	2	Disagree / No / Voluntary	239
	3	Neutral / Not sure	135
	4	Agree / Yes, Sometimes / Mandatory / Yes, but up to 10% higher / Yes (Q5)	308
	5	Strongly Agree / Yes, Always / Yes (Q4)	52
	Count (n)	**Total**	**750**
Hypothesis 5	1	Strongly Disagree	53
	2	Disagree	172
	3	Neutral	147
	4	Agree	187
	5	Strongly Agree	41
	Count (n)	**Total**	**600**

Table 10: Dataset per hypothesis for hypothesis testing

Using the sample data in Table 10 and the formulae indicated earlier in the methodology, the t-tests and hypothesis tests for the five hypotheses are performed in Microsoft Excel. The outcome and values off all tests are described in Table 11 below.

	Hypothesis Test 1	Hypothesis Test 2	Hypothesis Test 3	Hypothesis Test 4	Hypothesis Test 5
Null hypothesis (H0)	$\mu <= 3.00$	$\mu <= 3.00$	$\mu <= 3.00$	$\mu <= 3.00$	$\mu <= 3.00$
Alternative hypothesis (H1)	$\mu > 3.00$	$\mu > 3.00$	$\mu > 3.00$	$\mu > 3.00$	$\mu > 3.00$
Mean ($\overline{x}$)	3.58	3.31	3.70	3.19	2.99
Standard deviation (σ)	1.08	1.16	1.00	1.03	1.11
Count (n)	750.00	600.00	450.00	750.00	600.00
Standard error of mean ($\sigma\overline{x}$)	0.04	0.05	0.05	0.04	0.05
Degrees of freedom (df)	749.00	599.00	449.00	749.00	599.00
Hypothesized mean (μ)	3.00	3.00	3.00	3.00	3.00
Significance level (α)	0.05	0.05	0.05	0.05	0.05
t Critical one-tail	-1.65	-1.65	-1.65	-1.65	-1.65
t-Statistic	14.78	6.60	14.75	5.01	-0.33
p-Value	6.90E-44	4.45E-11	9.82E-41	3.48E-07	0.37
Reject H0 if t-statistic > t critical	**Rejected**	**Rejected**	**Rejected**	**Rejected**	**Rejected**
Reject H0 if p-value < α	**Rejected**	**Rejected**	**Rejected**	**Rejected**	**Failed to reject**

Table 11: t-Test and hypothesis test corresponding to the five null hypotheses

As seen in the t-tests above (Table 11), the t-statistical values of all five hypothesis tests are higher than the respective t-critical values. According to the data analysis methodology described earlier (e.g. Brownlee, 2020), this means that each null hypothesis can be rejected and that relationships between variables are present. However, the methodology also describes the use of the p-value to reject or fail to reject a hypothesis. Looking at these values, it is noted that not every null hypothesis can be rejected, as the p-value of H5 is higher than significance level. Because of this difference, it can be concluded that relationships are present between variables, but also that this relationship is not significant. Therefore, null hypothesis 5 cannot be rejected. An overview of the five hypothesis tests performed in this chapter is provided in Table 12 below.

Conclusion of Hypotheses Testing	
H1	European airline consumers think that there is a significant relationship between increasing airline emissions and rising global average temperatures
H2	There is a significant relationship between environmental ticket-booking factors, such as flight taxes and ecolabels, and the behavioural intention of consumers to book a flight ticket
H3	There is a significant relationship between the behavioural intention of consumers to pay a higher price for a flight ticket and the use of biofuel instead of kerosene by airlines
H4	There is a significant relationship between the behavioural intention of consumers to pay a higher price for a flight ticket and the applicant of carbon offsetting options on flight tickets
H5	There is no relationship between growing environmental awareness among European airline consumers and the image & quality perception of airlines

Table 12: Conclusion from five hypothesis tests

5.0. Discussion

The findings of the research are discussed by the key findings from the survey results and the hypotheses testing and are compared with the literature review. Both the questionnaire and the research hypotheses are structured according to the five main subjects. In accordance with these subjects, five key statements are made, starting with environmental awareness among consumers in the European airline industry.

Environmental awareness is present among airline consumers

Hypotheses testing statistically proved that European airline consumers think that increasing airline emissions correlate to rising global average temperatures. Researchers (Owen, Lee, & Lim, 2010; Graver, Zhang, & Rutherford, 2019) earlier indicated that CO_2 emissions from airlines increasingly contribute to rising global average temperatures. The questionnaire, moreover, showed that 83 per cent of the respondents (strongly) agree that carbon emissions contribute to rising global average temperatures and 79 per cent (strongly) agrees with the contribution of airlines. A vast majority of the respondents (82 per cent), on the other hand, also (strongly) agrees that environmental awareness is growing among the public, which confirms statements by various authors (Yoo, Divita, & Kim, 2013; Han, Yu, & Kim, 2019). However, less than 50 per cent of the sample believes that airlines share this environmental awareness and only few people (13 per cent) think that sustainable development in aviation is progressing fast enough. These findings confirm statements (Gössling & Peeters, 2009) that the aviation industry does not seem to recognise a sense of environmental importance. However, as 33 and respectively 30 per cent of the people remained neutral, it is possible that the sample did not have enough knowledge to make an informed decision (TalentMap, 2020), which could be solved by environmental education (Zsóka, Szerényi, Széchy, & Kocsis, 2013).

Sustainable ticket booking factors influence behavioural intentions

Applying environmental awareness to airline products, relationships between environmental ticket booking factors and consumer behavioural intention to purchase tickets were statistically demonstrated. However, the level of influence of sustainability is debatable. Although it was described that green image has a positive effect on consumers' attitudes (Hwang & Lyu, 2020), the questionnaire showed that only 22 respondents marked sustainability in general as an important factor to book a ticket. To put this in contrast, 138 people marked price as an important influencer, followed by flight time. Although price seems to be rated much more highly, as reported earlier (e.g. Dolnicar et al. , 2011), 69 per cent of the respondents did also (strongly) agree that sustainability by airlines does have an influence. On the other hand, people are not sure about the application of flight taxes on tickets. Although reports state that taxes can decrease aviation emissions (e.g. Krenek & Schratzenstaller, 2016), people largely answered 'not sure' when provided a comparable statement and only 52 per cent think that taxes should be implemented. Although people are unsure and might need additional environmental education (Zsóka et al. , 2013), this again shows that people share an environmental concern and want aviation to take actions. However, it also shows that the level of influence of sustainability on customers is different per specific ticket booking factor.

Airline consumers value the possibility of biofuel use by airlines
The environmental awareness among the public is also resembled in relation to the use of Sustainable Aviation Fuel by airlines. According to the corresponding hypothesis test, relationships are present between the intention to pay a higher ticket price and the use of SAF. Reports agree that SAF can emit up to 80 per cent less CO_2 than kerosene (SkyNRG, 2020), which is a statistic that the respondents (63 per cent) were unaware of. In relation to the earlier indicated environmental concern of the sample, 81 per cent of the people think that alternative fuel should be used more often by airlines. However, the fact that respondents believe that airlines should use SAF more often does not mean that the use of biofuel positively influences consumers' decision to book a ticket. Zeng et al. (2019), describe that environmental awareness by companies can be perceived by consumers as taking environmental responsibility. The very high benefits of 80 per cent less CO_2 in combination with a growing public environmental concern could trigger the valuation of SAF-use as a necessity rather than it being an amenity.

Nevertheless, people are still willing to pay higher prices to compensate for the high cost of production of SAF (e.g. International Energy Agency, 2019). Kumar, Garg and Makkar (2012) indicated that consumers are willing to pay up to 20 per cent higher prices for green products, but only 26 per cent of the respondents agree with that statement. A majority of 54 per cent states to be willing to pay higher ticket prices, but only up to 10 per cent higher. Here, it shows that the environmental concern of the people is weighed up against the earlier measured differences in importance in favour of ticket price against sustainability.

Airline consumers in theory value transparent carbon offsetting
Carbon offsetting practices by airlines are a method to decrease emissions in relation comparable to the use of SAF. However, the respondents are less convinced of the current effectiveness of the method. Statistically, tests described relationships between the intention to pay higher ticket prices for a ticket with carbon offsetting, but this is dependent on different aspects. First of all, people largely indicated to not use carbon offsetting options during ticket booking and also do not think that it is an effective method to decrease emissions, which was also described in the literature review earlier by Heikkinen (2020).

However, Heikkinen (2020) also elaborated that customers would be willing to pay higher ticket prices to offset the carbon of a flight. The survey results showed that 49 per cent of the respondents are willing to pay higher prices up to 10 per cent higher and that 20 per cent would be willing to pay over 10 per cent higher prices. Once again, this statistic resembles the environmental awareness and concern that was noted earlier. Moreover, as a solution to the current low number of customers that use carbon offsetting, 69 per cent of the people stated to be more willing to pay for carbon offsetting options if it is known how the payments are used by airlines. This way, the consumers' scepticism towards airlines' investment in environmental sustainability, as reported by Gössling and Peeters (2009), can be decreased.

Environmental Awareness does not significantly alter image and quality
The last research hypothesis proved to be interesting, as tests could not find significant relationships between environmental awareness and the image and quality perception of airlines, although it did state that relationships could be present. First, the differences in importance to consumers between price and sustainability were noted again, as 48 per cent of the respondents stated that customer satisfaction is affected by sustainability and 90 per cent indicated that this was influenced by price. Although Park et al. (2015) described that satisfaction is affected by (the lack of) environmental sustainability, only a minority of the respondents agrees with that statement. The latter could also relate to consumers' scepticism towards airlines, as only 19 per cent of the people (strongly) agree to believe airlines that claim to be CO_2 neutral.

The image of airlines, on the other hand, is affected more heavily by environmental awareness than by customer satisfaction. Almost 60 per cent of the people think the image of aviation is decreased by environmental awareness, which relates to the demonstrated growing public environmental concern. However, the image is not as bad that the majority of the people associates flying with a sense of guilt or "flight shame" (Hasberg, 2019). Only 16 per cent of the respondents agree to feel a sense of guilt, with a fast majority disagreeing with the statement. The results of this questionnaire subject do not elaborate that both airline image and quality perception are decreased by environmental awareness. Respondents are found to primarily believe that image is decreased by awareness, but that customer satisfaction is mainly affected by factors such as price. Therefore, as only one of two aspects of the hypothesis could be proven, significant relationships could not be determined by the statistical test.

6.0. Conclusion

Respondents were provided with a questionnaire that was structured according to five subjects – environmental awareness, ticket booking, SAF, carbon offsetting and consumer satisfaction & airline image – corresponding to the five research hypotheses. The relationships between variables in the five research hypotheses are tested using a statistical t-test. It is concluded that environmental awareness is present among airline consumers. A vast majority of the respondents of the questionnaire (strongly) agree that emissions cause global average temperatures to increase (83 per cent) and that airlines are contributing to this problem (79 per cent). Moreover, people do not agree that airlines understand the seriousness of climate change, which is in line with statements by Gössling and Peeters (2009).

Furthermore, environmental ticket booking factors are regarded by respondents to influence their ticket booking behaviour, but it is not the most important factor. For instance, 92 per cent of the sample indicate price as an important factor, with only 15 per cent marking sustainability. Therefore, it is concluded that there is a relationship between sustainability and consumer ticket booking behaviour, but also that other factors are more important to consumers. The level of significance, however, is based on the type of sustainable factor, as the research found that consumers prefer the application of ecolabels over fight ticket taxes.

The results to the hypotheses related to SAF and carbon offsetting showed similarities in the results from the questionnaire. After receiving information on the environmental benefits of SAF, a majority of the respondents (strongly) agreed that airlines should use biofuel more often and that they would pay a higher ticket price for the use of SAF. Carbon offsetting, on the other hand, received a similar response regarding additional payments. However, respondents also largely stated to not use and not believe in the effectiveness of carbon offsetting. The latter form of trust can be increased, as noted by the people, if airlines are clearer about the use of additional consumer payments. Moreover, the majority also indicated to be willing to pay a maximum additional amount of 10 per cent of the ticket price, which is different than the 20 per cent as described by the literature. It is concluded that a present environmental awareness, as determined earlier, is resembled in the willingness of consumers to pay a higher ticket price if airlines use SAF or carbon offsetting, but that the level of willingness is depending on the level of trust in airlines to use additional customer payments for the development of environmental sustainability.

The only relationship from the research hypotheses that is not proven as significant, is the relationship between environmental awareness and the image and quality of airlines. The literature elaborated that the image of the industry is decreasing because of environmental awareness, of which 59 per cent of the respondents (strongly) agree with. However, consumers also indicate that price is more important to influence customer satisfaction than environmental sustainability, which was also found in earlier results. Moreover, it is also indicated that people do not trust airlines that claim to take environmental responsibility. These results conclude that the relationship between both the image and quality perception of airlines and environmental awareness is not significant, as respondents value price over sustainability in terms of customer satisfaction. However, it is also concluded that image itself is affected by environmental awareness.

6.1. Conclusion Overview

The objective of this paper was to identify and assess relationships between consumer behaviour and environmental awareness & sustainability from the perspective of European airline consumers using quantitative research. An overview of the conclusion to the research objective is provided in Table 13 below.

Conclusion Overview
Environmental *concern* and awareness are increasingly present among European Airline consumers.
The price of a ticket is the most important factor to influence consumer behaviour, although consumers do value environmental sustainability by airlines.
Consumers are willing to pay higher ticket prices for the use of sustainable aviation fuel and carbon offsetting options by airlines, although the level of willingness is affected by the level of trust in an airline.
The image of airlines is decreasing due to increasing environmental awareness, but this does not significantly influence customer satisfaction regarding airlines.

Table 13: Overview and summary of the research conclusion

6.2. Recommendations to Aviation Industry

In the research, three components influence the conclusion: environmental awareness, environmental sustainability and ticket price. A combination of the three can be used to provide solutions to the industry regarding sustainable development. It is recommended to the industry to install a flight ticket tax on all flights from and to Europe to a maximum of 10 per cent of the ticket price. As the research found, the production of SAF is more expensive than kerosene. The income from ticket taxes can be used by the European Commission to subsidize sustainable fuel production, which decreases the utilization costs for airlines. Therefore, it becomes more economically attractive for airlines to use, increasing SAF-utilization and decreasing aviation emissions. Figure 3 below shows an ongoing development cycle corresponding to this recommendation.

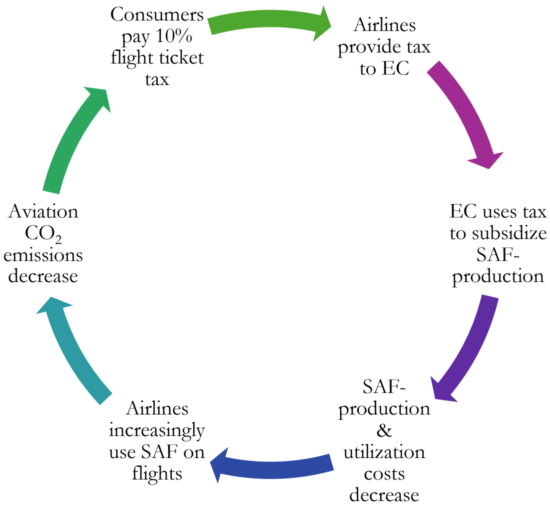

Figure 2: Recommended SAF-development cycle in aviation

However, to make the additional payment attractive to the consumer, it is recommended to provide more clarity on the sustainable developments that are made in the industry. People need to be informed on the use of the additional ticket tax, and the EC needs to provide information on the development of SAF-use.

References

Abdul Rashid, N.R., Abdul Rahman, N.I. & Khalid, S.A. (2014) 'Environmental corporate social responsibility (ECSR) as a strategic marketing initiative', *Procedia – Social and Behavioural Sciences*, 130, pp. 499–508.

Adom, D., Hussein, E.K. & Agyem, J.A. (2018) 'Theoretical and conceptual framework: Mandatory ingredients of quality research', *International Journal of Scientific Research*, 7, pp. 438–441.

Angelovska, J., Bilic Sotiroska, S. & Angelovska, N. (2012) 'The impact of environmental concern and awareness on consumer behaviour', *Journal of International Environmental Application & Science*, 7, pp. 406–416.

ATAG (2020) *Sustainable aviation fuel.* Available at: https://aviationbenefits.org/environmental-efficiency/climate-action/sustainable-aviation-fuel (Accessed 6 July 2020).

Bhate, S. (2001) 'One world, one environment, one vision: Are we close to achieving this? An exploratory study of consumer environmental behaviour across three countries', *Journal of Consumer Behaviour*, 2, pp. 169–184.

Bittner, A., Wallace, T.E. & Zhao, X. (2015) 'Field to flight: A techno-economic analysis of the corn stover to aviation biofuels supply chain', *Biofuels, Bioproducts and Biorefining*, 9, pp. 201–210.

Blaikie, N. (2003) *Analysing Quantitative Data*. London: SAGE Publications.

Brownlee, J. (2020) *A gentle introduction to statistical hypothesis testing*. Available at: https://machinelearningmastery.com/statistical-hypothesis-tests/(Accessed 10 April 2020).

Brueckner, J.K. & Abreu, C. (2017) 'Airline fuel usage and carbon emissions: Determining factors', *Journal of Air Transport Management*, 62, pp. 10–17.

Burrell, G. & Morgan, G. (2016) *Sociological Paradigms and Organisational Analysis*. Abingdon: Routledge.

Chen, F-Y., Chang, Y-H. & Lin, Y-H. (2012) 'Customer perceptions of airline social responsibility and its effect on loyalty', *Journal of Air Transport Management*, 20, pp. 49–51.

Chiaramonti, D., Prussi, M., Buffi, M. & Tacconi, D. (2014) 'Sustainable bio-kerosene: Process routes and industrial demonstration activities in aviation biofuels', *Applied Energy*, 136, pp. 767–774.

Cohen, S. (2015) *The growing level of environmental awareness*. Available at: https://www.huffpost.com/entry/the-growing-level-of-envi_b_6390054 (Accessed 28 February 2015).

Dolnicar, S., Grabler, K., Grün, B. & Kulnig, A. (2011) 'Key drivers of airline loyalty', *Tourism Management*, 32, pp. 1020–1026.

EASA Europa (2020) *Sustainable aviation fuels*. Available at: https://www.easa.europa.eu/eaer/climate-change/sustainable-aviation-fuels (Accessed 6 July 2020).

European Commission (2020a) *Biofuels in aviation – greening the skies*. Available at: https://setis.ec.europa.eu/publications/setis-magazine/bioenergy/biofuels-aviation-%E2%80%93-greening-skies (Accessed 3 July 2020).

European Commission (2020b) *Inception impact assessment*. Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12494. (Accessed 28 August 2020).

European Commission (2020c) *Reducing emissions from aviation*. Available at: https://ec.europa.eu/clima/policies/transport/aviation_en (Accessed 6 October 2020).

European Technology and Innovation Platform (2020) *HVO/HEFA*. Available at: https://www.etipbioenergy.eu/value-chains/products-end-use/products/hvo-hefa (Accessed 8 July 2020).

FitzGerald, J. & Tol, R.S. (2007) *Airline emissions of carbon dioxide in the European trading system*. Dublin: ESRI.

Francis, G. & Jakicic, V. (2022) 'Equivalent statistics for a one-sample t-test', *Behaviour Research Methods*, 55, pp. 77–84.

Frost, J. (2020) *How to analyse Likert scale data*. Available at: https://statisticsbyjim.com/hypothesis-testing/analyze-likert-scale-data/ (Accessed 25 September 2020).

Gadenne, D.L., Kennedy, J. & McKeiver, C. (2009) 'An empirical study of environmental awareness and practices in SMEs', *Journal of Business Ethics*, 84, pp. 45–63.

Gastwirth, J.L. & Rubin, H. (1971) 'Effect of dependence on the level of some one-sample tests', *Journal of the American Statistical Association*, 66, pp. 816–820.

Glen, S. (2020a) *Hypothesis testing*. Available at: https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/(Accessed 24 September 2020).

Glen, S. (2020b) *P-value in statistical hypothesis tests*. Available at: https://www.statisticshowto.com/p-value/ (Accessed 25 September 2020).

Glen, S. (2020c) *T critical value: Easy definition*. Available at: https://www.statisticshowto.com/t-critical-value/ (Accessed 25 September 2020).

Gössling, S. & Peeters, P. (2009) 'An analysis of industry discourses on tourism, air travel and the environment', *Journal of Sustainable Tourism*, pp. 402–417.

Graham, B. & Shaw, J. (2008) 'Low-cost airlines in Europe: Reconciling liberalisation and sustainability', *Geoforum*, 39, pp. 1439–1451.

Graver, B., Zhang, K. & Rutherford, D. (2019) *CO_2 emissions from commercial aviation, 2018*. Washington DC: ICCT.

Grimme, W. (2008) 'Measuring the long-term sustainability of air transport', *European Transport Conference*, pp. 1–21.

Han, H., Yu, J. & Kim, W. (2019) 'Environmental corporate social responsibility and strategies to boost airline image and loyalty', *Journal of Travel & Tourism Marketing*, 35, pp. 371–383.

Hasberg, K.S. (2019) *Research note: Flight shame*. Aalborg: Aalborg University.

Heikkinen, T.T. (2020) *The effectiveness of carbon offsets in increasing the environmental image of airlines*. Espoo: Aalto University.

Hwang, J. & Lyu, S.O. (2020) 'Relationships among green image, consumer attitudes, desire and customer citizenship behaviour in the airline industry', *International Journal of Sustainable Transportation*, 14, pp. 437–447.

IATA (2019) *Sustainable aviation fuels fact sheet*. Montreal: International Air Transport Association.

International Energy Agency (2019) *Are aviation biofuels ready for take-off?* Available at: https://www.iea.org/commentaries/are-aviation-biofuels-ready-for-take-off (iea.org in Bing) (Accessed 18 March 2019).

IPCC (2020) *History of the IPCC*. Available at: https://www.ipcc.ch/about/history/ (Accessed 6 October 2020).

Johnson, D.H. (1999) 'The insignificance of statistical significance testing', *The Journal of Wildlife Management*, 63, pp. 763–772.

Joseph, O.O. (2020) 'Pro-environmental consumer behavior: A critical review of literature', *International Journal of Business and Management*, 15, pp. 1–15.

Karaman, A.S. & Akman, E. (2018) 'Taking-off corporate social responsibility programs: An AHP application in the airline industry', *Journal of Air Transport Management*, 68, pp. 187–197.

Kent State University (2020) *One-sample t-test*. Available at: https://libguides.library.kent.edu/SPSS/OneSampletTest (libguides.library.kent.edu in Bing) (Accessed 22 September 2020).

Kenton, W. (2019) *Descriptive statistics*. Available at: https://www.investopedia.com/terms/d/descriptive_statistics.asp (investopedia.com in Bing) (Accessed 27 June 2019).

Kim, Y., Lee, J. & Ahn, J. (2019) 'Innovation towards sustainable technologies: A socio-technical perspective on accelerating transition to aviation biofuel', *Technological Forecasting & Social Change*, 145, pp. 317–329.

Krenek, A. & Schratzenstaller, M. (2016) *Sustainability-oriented EU taxes: The example of a European carbon-based flight ticket tax*. FairTax WP-Series No. 1.

Kumar, S., Garg, R. & Makkar, A. (2012) 'Consumer awareness and perception towards green products', *International Journal of Marketing & Business Communication*, 1, pp. 35–44.

Lewis-Beck, M.S., Bryman, A. & Liao, T.F. (2004) *The SAGE Encyclopedia of Social Science Research Methods*. New York: SAGE.

Klöwer, M. (2021) 'Quantifying aviation's contribution to global warming', *Environmental Research Letters*, Article 104027.

McDonald, S. & Oates, C.J. (2006) 'Sustainability: Consumer perceptions and marketing strategies', *Business Strategy and the Environment*, 15, pp. 157–170.

NCSS (2020) *One-sample t-test*. Available at: https://ncss-wpengine.netdna-ssl.com/.(Accessed 25 September 2020).

Neste Oil (2020) *Neste's role in sustainable aviation*. Available at: https://www.neste.com/companies/products/aviation/neste-my-renewable-jet-fuel (Accessed 8 July 2020).

Newman, I. (2000) *A conceptualization of mixed methods*. AERA Annual Meeting, New Orleans.

Park, E., Lee, S., Kwon, S.J. & del Pobil, A.P. (2015) 'Determinants of behavioural intention to use South Korean airline services', *Sustainability*, 7, pp. 12106–12121.

Peeters, P. et al. (2016) 'Are technology myths stalling aviation climate policy?', *Transportation Research Part D*, 44, pp. 30–42.

Phylactou, P., Papadatou-Pastou, M. & Kostantinou, N. (2025) 'The Bayesian one-sample t-test supersedes correlation analysis', *European Journal of Psychology Open*, 84(1), pp. 1–12.

Ponto, J. (2015) 'Understanding and evaluating survey research', *Journal of the Advanced Practitioner in Oncology*, 6, pp. 168–171.

Prussi, M., O'Connell, A. & Lonza, L. (2019) 'Analysis of current aviation biofuel technical production potential in EU28', *Biomass and Bioenergy*, 130, pp. 1–8.

Rajaguru, R. (2016) 'Role of value for money and service quality on behavioural intention', *Journal of Air Transport Management*, 53, pp. 114–122.

Rupcic, L. et al. (2023) 'Environmental impacts in the civil aviation sector', *Transportation Research Part D*, 119.

Salvioni, D.M., Gennari, F. & Bosetti, L. (2016) 'Sustainability and convergence: The future of corporate governance systems?', *Sustainability*, 8, pp. 1–25.

Saunders, M.N., Lewis, P. & Thornhill, A. (2019) *Research Methods for Business Students*. Harlow: Pearson.

Singleton, R.A. & Straits, B.C. (2009) *Approaches to Social Research*. New York: Oxford University Press.

TalentMap (2020) *What to do with neutral employee survey responses*. Available at: https://talentmap.com/neutral-responses-employee-surveys/.(Accessed 10 September 2020).

The Pennsylvania State University (2020) *Hypothesis testing (P-value approach)*. Available at: https://online.stat.psu.edu/. (Accessed 28 September 2020).

The World Bank Group (2020) *Air transport, passengers carried*. Available at: https://data.worldbank.org/indicator/IS.AIR.PSGR. (Accessed 6 October 2020).

UNFCCC (2020) *History of the Convention*. Available at: https://unfccc.int/process/the-convention/history-of-the-convention. (Accessed 6 October 2020).

Walker, S. & Cook, M. (2009) 'The contested concept of sustainable aviation', *Sustainable Development*, 17, pp. 378–390.

Watson, R.R. (2010) *Handbook of Disease Burdens and Quality of Life Measures*. New York: Springer.

Wilson, J. (2014) *Essentials of Business Research*. London: SAGE Publications.

Yang, H. & Chen, W. (2018) 'Retailer-driven carbon emission abatement with consumer environmental awareness', *Omega*, 78, pp. 179–191.

Yoo, J-J., Divita, L. & Kim, H-Y. (2013) 'Environmental awareness on bamboo product purchase intentions', *International Journal of Fashion Design, Technology and Education*, 6, pp. 27–34.

Zeng, S., Qin, Y. & Zeng, G. (2019) 'Impact of corporate environmental responsibility on investment efficiency', *Sustainability*, 11, pp. 1–21.

Zsóka, Á., Szerényi, Z.M., Széchy, A. & Kocsis, T. (2013) 'Greening due to environmental education?', *Journal of Cleaner Production*, 48, pp. 126–138.

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 33-44
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1783

Preserving Cognitive Ownership of Academic Writing in Higher Education: A Sustainable Hybrid Pedagogical Framework for Reasoning-Centred Artificial Intelligence Integration

[1]Sunika Naz, [2]Stella Sardar – [1,2]Arden University
[3]Imad Yasir Nawaz- Northumbria University, London

Abstract
This study explores how distinctive types of generative artificial intelligence (AI) practice, augmentation, co-construction, and replacement shape students' reasoning skills and sense of cognitive ownership in higher education (HE) academic writing. This research also responds to growing humanitarian concerns about the erosion of student commitment, the undermining of autonomy, and ethical learning in HE. To address this core gap, an explanatory sequential mixed-methods design was employed. Data were collected from 412 UK HE students, complemented with in-depth interviews from 24 participants. Quantitative modelling showed that augmentation strengthens reasoning through reflective engagement, co-construction yields mixed cognitive outcomes, and replacement significantly weakens ownership and efficacy. Qualitative findings revealed subsistent experiences behind these practices: some students articulated no ethical harm by AI-supported reflection, while others exhibited a quiet disarticulation of their self-learning skills. Incorporating these insights, this study proposed the Hybrid Human–AI Reasoning Integrity Model (HHARIM), a sustainable pedagogical framework in HE that centres human reasoning in ethical AI use. The recommended model also highlights cognitive ownership as an essential element and outlines a robust framework for responsible AI use to safeguard learning, ethics, and autonomy in HE. This study contributes theoretically by offering HHARIM as a framework for effectively embedding AI, thereby upholding ethical, sustainable, and human-centred learning. Ultimately, the implications of this proposed model will influence HE systems to encourage sustainable AI pedagogical practices that reinforce academic writing rather than compromise students' learning efficacy.

Keywords: Artificial Intelligence; Cognitive ownership; Sustainable Pedagogical; Higher Education.

Wordcount: 234

1.0 Introduction

The rapid evolution of generative Artificial Intelligence (AI) has initiated profound transformations in higher education (HE), fundamentally impacting traditional methods of student research, brainstorming, and academic writing. Since the emergence of large language models (LLMs) such as ChatGPT, Gemini, Claude, Grammarly and Copilot, these AI systems have become increasingly embedded in students' everyday learning practices, reshaping core academic behaviours related to critical thinking, assignment writing, and intellectual agency (Kasneci et al., 2023; Luckin, 2024; Zawacki-Richter et al., 2019). While universities have rigorous restrictions on plagiarism, authorship, and academic misconduct (Bretag, 2019; Dawson, 2020; Eaton, 2023), far less attention has been given to how AI influences cognitive ownership; the extent to which students retain responsibility and authorship over reasoning processes that assignment writing is designed to develop (UNESCO, 2023).

Academic writing has long served as a fundamental pedagogical tool for assessing knowledge in relevant areas, cultivating critical reasoning, and disciplinary identity (Bean, 2011; Wingate, 2012). Through processes of argument construction, evidence evaluation, explanation, and reflection, students develop the cognitive foundations essential to higher-order thinking (Andrews, 2010). However, the increasing availability of generative AI now challenges these assessments. AI's capacity to produce structured arguments, coherent explanations, and discipline-specific interpretations raises critical questions regarding what remains "human" in the reasoning demonstrated through written assignments (Selwyn, 2023; Williamson & Piattoeva, 2023). These concerns extend beyond textual originality to encompass deeper issues of human reasoning integrity, cognitive displacement, and intellectual authorship.

Emerging research suggests that AI can both enhance and abolish students' reasoning. Again, AI may provide scaffolding that strengthens metacognition, prompts reflection, and supports HE learners who struggle with structure or language (Holmes et al., 2022; Kasneci et al., 2023). However, over-reliance on AI-generated academic writing can lead to cognitive offloading, in which essential learning tasks are generated without human involvement, reducing opportunities for students to develop academic writing and critical thinking skills (Fischer et al., 2020; Yang, 2022). Existing literature and institutional policy, however, remain ambiguous and provide limited empirical clarity on how different modes of AI should be used, such as augmentation, co-construction, or replacement, to shape students' learning processes.

This significant gap in evidence is increasingly problematic as generative AI becomes normalised within higher education. Students often lack sufficient knowledge of the ethical boundaries between appropriate and inappropriate AI assistance (Cotton et al., 2023; Searle et al., 2024), and institutional guidelines rarely address the cognitive implications of AI-supported writing. UNESCO (2023) has called for governance models that prioritise human productivity and ethical decision-making, yet practical, pedagogically grounded frameworks remain scarce. There is an urgent need to reconceptualise academic integrity around reasoning integrity, the preservation of human-led cognitive processes within AI-supported writing, rather than focusing solely on textual originality.

To address these gaps, this study investigates: To what extent do different modes of AI use influence students' reasoning skills and cognitive authorship in academic writing, and how can HE integrate AI ethically while preserving human intellectual agency?

Drawing on a mixed-methods design, the study combines structural equation modelling (SEM) with qualitative thematic analysis to examine relationships between AI use, human reasoning performance, and authorship perceptions. In doing so, it foregrounds the cognitive dimensions of academic integrity and introduces the Hybrid Human–AI Reasoning Integrity Model (HHARIM), a framework designed to support ethical, reasoning-centred AI integration in higher education. By addressing the cognitive, ethical, and pedagogical implications of AI-supported writing, this study responds to a growing need for evidence-based strategies that leverage AI's educational potential while safeguarding students' knowledge and skill development.

Generally, this research aims to contribute, both practically and theoretically, by proposing HHARIM, a sustainable, pedagogical and governance-oriented model that encourages ethical, sustainable, and human-centred AI adoption in HE.

2.0 Literature Review

Academic writing has long been acknowledged as a core approach through which higher education assesses students' critical reasoning, skills, academic knowledge and disciplinary identity (Lee et al., 2024). Over the past two decades, research has emphasised that writing is not merely a communicative output but a complex cognitive process that enables interpretation, argumentation, and reflective judgement (Andrews, 2010; Lea & Street, 1998; Wingate, 2012). Through the iterative structure of drafting, reviewing and revising, students internalise the reasoning skills that underpin advanced academic competence (Bean, 2011). Yet, as generative AI tools accelerate rapidly, this pedagogical function is undergoing significant disruption.

Generative AI and Cognitive Transformation in HE

Generative AI systems particularly large language models (LLMs) are now widely used across HE for brainstorming, summarising texts, revising drafts, and transcribing conceptual explanations (Kasneci et al., 2023; Zawacki-Richter et al., 2019). Scholars have recognised that AI can provide linguistic scaffolding, personalised feedback, and cognitive support, potentially enhancing learning, especially for students facing linguistic or structural writing challenges (Holmes et al., 2022; Dwivedi et al., 2024). Moreover, research grounded in digital learning frameworks suggests that AI-enabled tools may improve learning efficacy, reduce writing anxiety, and broaden access to academic support (Thompson & Lee, 2012; Weller, 2020).

However, the transformative capacity of AI raises pedagogical and epistemic concerns. AI-generated text can resemble high-quality academic writing without requiring the underlying reasoning effort traditionally associated with scholarly work (Selwyn, 2023; Williamson & Piattoeva, 2023). This creates risks of cognitive offloading, where essential tasks such as conceptualisation, evidence evaluation and argument construction are transferred from the student to the AI system (Fischer et al., 2020; Sparrow & Chatman, 2019). As a result, academic integrity concerns have expanded beyond issues of plagiarism and authorship to include the erosion of critical thinking skills and intellectual autonomy (Bretag, 2019; Eaton, 2023).

Academic Integrity Beyond Plagiarism: The Rise of Cognitive Integrity

Between 2000 and 2020, academic integrity discourse was dominated by concerns over cheating, plagiarism detection, and text-based misconduct (Dawson, 2020; Macfarlane et al., 2014). While these concerns remain relevant, scholars now argue that generative AI necessitates a broader conceptualisation of integrity, one that includes cognitive effort, reasoning transparency, and the human ownership of intellectual processes (Eaton, 2023; Harper et al., 2021). Moreover, UNESCO's (2023) global guidance echoes this shift by emphasising human agency as a non-negotiable foundation for AI use in HE.

This emerging discourse suggests that traditional plagiarism frameworks are insufficient for addressing AI's cognitive implications. When AI produces arguments, explanations, or conceptual connections on behalf of students, the issue is not simply one of textual authorship but of diminished cognitive ownership, a student's sense of control and responsibility over the reasoning embedded in their coursework (Luckin, 2024). Loss of cognitive ownership undermines the very purpose of academic writing as a process for developing critical thinking (Lea & Street, 1998; Wingate, 2012).

AI-Augmented, Co-Constructed, and AI-Generated Reasoning

Recent literature distinguishes between three primary modes of student–AI interaction: augmentation, co-construction, and replacement. Augmentation involves using AI as a scaffold to support human reasoning—for example, brainstorming ideas, generating outlines, clarifying concepts, or prompting metacognitive reflection. Existing Research shows that augmentation can enhance reasoning by stimulating cognitive engagement and enabling deeper reflection (Kasneci et al., 2023; Holmes et al., 2022).

Co-construction reflects iterative collaboration between human reasoning and AI-generated suggestions. While this mode can encourage active decision-making and comparative evaluation, scholars warn that constant back-and-forth can dilute cognitive effort, as learners may over-rely on AI prompts (Yang, 2022; Popenici & Kerr, 2017).

Replacement describes scenarios in which AI performs core reasoning tasks, such as interpreting data, generating arguments, or drafting paragraphs on behalf of students (Crompton & Burke, 2023). Studies repeatedly show that replacement undermines learning by reducing cognitive effort, weakening reasoning skills, and limiting the development of intellectual agency (Fischer et al., 2020; Selwyn, 2023; Williamson & Eynon, 2020).

Despite growing conceptual clarity, empirical research examining the relationship between AI use and human reasoning outcomes remains limited. Existing studies focus largely on user perceptions, ethical concerns, or plagiarism risks, rather than outlining how different patterns of AI engagement influence human reasoning performance or ownership (Cotton et al., 2023; Searle et al., 2024). This gap highlights the need for robust empirical evidence.

Cognitive Ownership as a Mediating Construct

While cognitive ownership is increasingly recognised in theoretical discussions, few studies have empirically measured it or examined its structural role in student learning (Bailey, 2023; Harper et al., 2021). The literature emphasises that high levels of ownership are associated with deeper cognitive engagement, greater metacognitive awareness, and stronger intellectual autonomy (Bandura, 2001; Zimmerman, 2002). Conversely, cognitive displacement through automation can weaken agency and long-term reasoning capability (Fischer et al., 2020). Recent theorists argue that AI's educational impact cannot be understood solely through performance outcomes; instead, it must be analysed through the lens of ownership and authorship of cognition (Luckin, 2024; Selwyn, 2023). This perspective positions cognitive ownership as a potential mechanism explaining why augmentation supports reasoning and why replacement undermines it.

However, the mediating role of cognitive ownership in AI-assisted writing remains underexplored empirically. The literature lacks mixed-methods studies that integrate quantitative measurement of reasoning with qualitative insights into students' lived cognitive experiences.

3.0 Ethical and Pedagogical Frameworks for AI Integration

The expanding presence of AI in education has prompted calls for governance frameworks that preserve human agency and support ethical decision-making (UNESCO, 2023; Eaton, 2023). Existing literature proposes that ethical AI use in HE should prioritise transparency, human-led reasoning, and pedagogical structures that discourage cognitive offloading (Holmes & Tuomi, 2022; Williamson & Eynon, 2020). Yet existing frameworks tend to focus on institutional policy rather than on the pedagogical or cognitive mechanisms that preserve the integrity of reasoning.

As Weller (2020) and Popenici and Kerr (2017) highlight, meaningful AI integration must include pedagogical safeguards such as reflective tasks, justification prompts, and metacognitive checkpoints to ensure students remain the primary originators of their thinking. Without these safeguards, AI risks transforming academic writing from a reasoning-centred practice into a productivity exercise detached from cognitive development. Additionally, UNESCO (2023) highlights the importance of placing human reasoning as the principle of creation. While AI can be an effective tool for proofreading, there are still only a few practical pedagogical models available.

After critically scrutinising the existing literature, the core emerging gap is identified: the absence of sustainable, pedagogically grounded models that pilot ethical, human-reasoning-centred AI integration in HE. This study addresses these gaps by empirically evaluating AI-use modes, modelling the role of cognitive ownership in reasoning performance, and proposing the Hybrid Human–AI Reasoning Integrity Model (HHARIM) as a framework for sustainable, ethical and human-centred AI adoption in higher education.

4.0 Conceptual Model Summary

The framework ascertains AI-use modes as the primary independent variable, human reasoning skills as the dependent variable, and cognitive ownership as the mediating mechanism that influences the relationship. Reasoning integrity frames the broader ethical context.

This conceptual grounding informed the following study hypotheses:

- H01: AI-use mode predicts reasoning skills.
- H02: Cognitive ownership mediates the relationship between AI-use mode and reasoning skills.
- H03: AI-use mode influences perceived academic integrity through a sequential pathway:

AI-use mode → cognitive ownership → human reasoning skills → integrity perceptions.

The conceptual logic underlying this study begins with the recognition that generative AI has become deeply embedded in academic writing, fundamentally altering how students approach reasoning, ownership, and knowledge-based work. As students interact with AI in different ways, two distinct modes of use emerge: augmentation and replacement, each carrying different cognitive implications. Augmentation supports students' thinking by prompting reflection, offering scaffolding, or generating examples, thereby reinforcing their cognitive ownership. In contrast, replacement allows AI to take over the reasoning process, producing arguments and explanations on the student's behalf, which risks diminishing the very reasoning skills academic writing is designed to develop. This presents a challenge to both reasoning integrity, the preservation of human-led cognitive processes, and cognitive ownership, the sense that students retain authorship over their knowledge construction. Because these risks arise not from the technology alone but from how it is used, effective responses require a combination of governance (clear rules, transparency, institutional oversight) and pedagogy (teaching strategies that foreground reasoning and metacognition). Together, these elements define safe, ethical, and academically sound boundaries for AI use. The Hybrid Human–AI Reasoning Integrity Model (HHARIM) embodies this logic by positioning human reasoning as the anchor of all AI-supported academic writing practices, ensuring that AI acts as a tool for knowledge gathering rather than a substitute for cognitive effort. In doing so, the framework promotes sustainable, ethical, and human-centred reasoning in academic practices in HE.

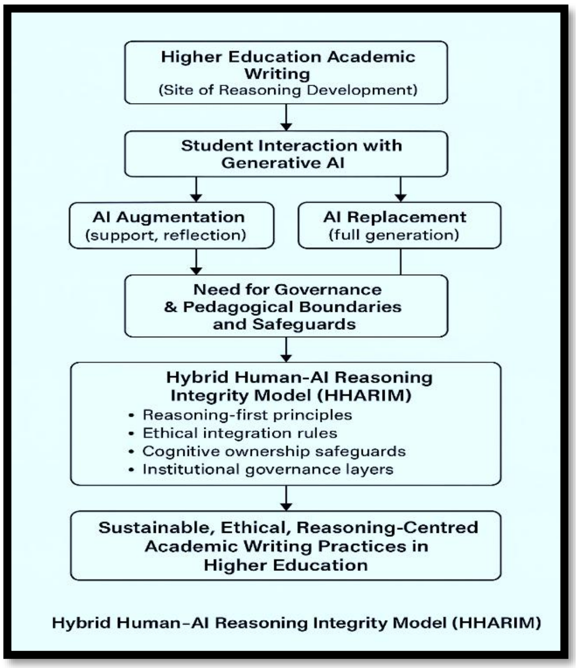

Figure 1: Proposed Conceptual Framework.

5.0. Methodology

This study adopted an explanatory sequential mixed-methods design, grounded in a pragmatic paradigm, to investigate the extent to which and how different generative AI-use modes influence students' reasoning and cognitive ownership. The quantitative phase involved 412 UK HE students recruited through voluntary sampling, using a survey measuring AI-use mode, reasoning skills (CRI), and a validated five-item cognitive ownership scale. Data were analysed using SPSS and AMOS SEM, enabling examination of group differences and mediating pathways. The qualitative phase purposively selected 24 diverse participants for semi-structured interviews exploring human-centred reasoning processes, ownership, and ethical boundaries in AI use. Thematic analysis, supported by NVivo, combined inductive and deductive coding, with credibility ensured through triangulation, peer debriefing, and an audit trail. Integration occurred through side-by-side comparison of themes and SEM pathways, informing the development of the Hybrid Human–AI Reasoning Integrity Model (HHARIM). Ethical approval and standard protocols were observed throughout.

The sample represented a broad mix of disciplines, academic years, and levels of AI familiarity. Participants provided information on their frequency and mode of AI use in academic writing, self-rated their reasoning ability, and completed the validated Critical Reasoning Inventory (CRI) (1–5 scale), which measures analytical thinking, argument construction, evidence evaluation, and reflective judgement.

• AI-Use Mode: Students selected statements corresponding to one of three theoretically derived categories—Augmentation, Co-construction, and Replacement—reflecting the degree to which AI supports, collaborates in, or substitutes for their reasoning.
• Reasoning Skills (CRI Score): Established measure assessing analytical and evaluative reasoning competencies.
• Perceived Cognitive Ownership: A five-item scale ($\alpha = .82$) developed for this study to assess the extent to which students believed their submitted writing represented their own thinking.
• Control Variables: Discipline (STEM vs. non-STEM), year of study, and prior academic writing experience.

Again, semi-structured interviews using maximum variation sampling based on AI-use intensity (high/low), discipline, and gender ensured that diverse user profiles were represented, particularly in terms of their reasoning approaches and experiences with AI-supported writing. Interviews explored:
• how students engaged with AI during their writing process,
• how they understood the boundary between their own reasoning and AI contributions,
• how they interpreted ownership of ideas and academic integrity.

Building on the literature that distinguishes among augmentation, co-construction, and replacement modes of AI use (Holmes et al., 2022; Selwyn, 2023) and their divergent effects on reasoning, and grounded in the HRF principle, which highlights human-led cognitive processes, this study tested three hypotheses. These hypotheses reflect mediation and sequential-mediation relationships that align with this study's variables and conceptual framework.

6.0. Findings

Consistent with the explanatory sequential design, the quantitative results established the structural relationships among key variables, and the qualitative analysis subsequently enriched and explained these patterns. The integration of these two strands offers a nuanced account of students' reasoning practices in the context of AI-supported writing. It directly informs the development of the Hybrid Human–AI Reasoning Integrity Model (HHARIM).

The quantitative phase (n = 412) revealed apparent differences in reasoning performance and cognitive ownership across the three AI-use modes. Augmentation was the most common mode (42%), followed by Co-construction (31%) and Replacement (27%). Mean CRI scores varied significantly: Augmentation users scored highest (3.73), Co-construction users scored moderately (3.39), and Replacement users scored lowest (2.94). ANOVA results confirmed significant differences across groups, $F(2,409) = 24.6$, $p < 0.001$, with Tukey tests revealing the hierarchy: Augmentation > Co-construction > Replacement. Cognitive ownership scores followed the same pattern, with augmentation users reporting the highest ownership ($M = 4.12$) and replacement users reporting the lowest ($M = 2.68$). Structural equation modelling ($\chi^2/df = 1.97$, CFI = 0.96, RMSEA = 0.045) demonstrated that AI-use mode significantly predicted reasoning skills ($\beta = 0.43$, $p < 0.001$) and perceived cognitive ownership ($\beta = 0.57$, $p < 0.001$). Reasoning skills also predicted ownership ($\beta = 0.51$, $p < 0.001$), supporting a sequential-mediation pathway in line with the study's hypotheses. Control variables were non-significant.

Figures 2 and 3 show the distribution of AI-use modes among students: Augmentation (42%), Co-construction (31%), Replacement (27%). Figure 3 demonstrates the mean CRI reasoning scores for each AI-use mode: Augmentation = 3.73, Co-construction = 3.39, Replacement = 2.94.

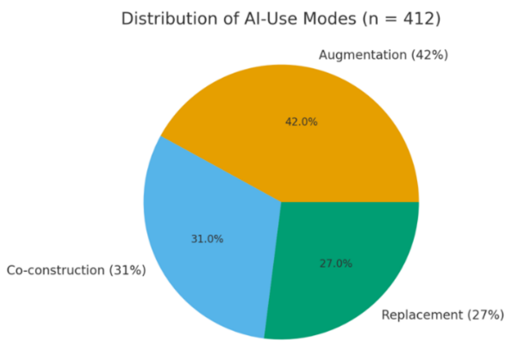

Figure 2: Distribution of AI-use modes (n=412).

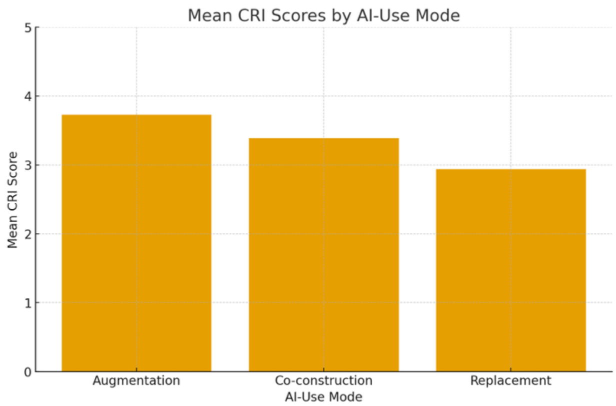

Figure 3: Mean CRI Scores by AI-use mode.

In addition to examining AI-use modes, the quantitative findings also revealed clear patterns in the specific AI tools students used for academic writing. As shown in Figure 4, students engage with a broad ecosystem of AI technologies, including both traditional text-enhancement tools and advanced generative AI systems. Grammarly (96 users) and ChatGPT (90 users) emerged as the most frequently used tools, followed by Microsoft Copilot (77 users), Google Gemini (75 users), and Quillbot (74 users). This highlights the pedagogical importance of understanding how different tools contribute to augmentation, co-construction, or replacement modes of AI use.

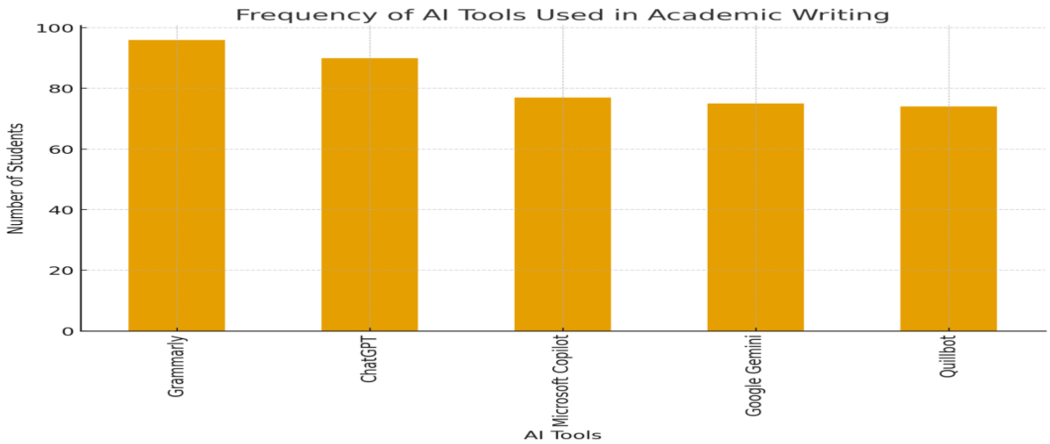

Figure 4: Frequency of AI tools used in academic writing

Reasoning Displacement (Explains Low Scores for Replacement Users)

Semi-structured interviews with 24 students elaborated and contextualised the statistical results. Four themes captured students lived experiences of AI-supported writing and revealed nuanced mechanisms behind the quantitative relationships. Students who used AI in a replacement mode consistently reported that AI "took over" the cognitive work. Many described submitting text they did not fully understand:
"It sounded good, but if someone asked me to explain it, I wouldn't know where to start." This theme helps explain the lower CRI scores and lower cognitive ownership observed among replacement users in the quantitative data. Participants explicitly linked reduced reasoning effort to reduced ownership, a pattern mirrored in the SEM pathway, which showed a negative association between replacement use and reasoning skills.

AI-Enabled Reflection (Explains Higher Scores for Augmentation Users)

Students who used AI augmentatively described AI as a reflective tool that supported idea generation, organisation, and metacognitive awareness: *"AI helps me notice gaps in my argument—it doesn't replace my thinking but makes me think more clearly."* This theme explains why Augmentation users showed the strongest CRI and ownership scores. It aligns with the SEM result, indicating a strong positive path from AI-use mode to reasoning skills when the use is supportive rather than substitutive.

Ethical Augmentation (Explains Strong Ownership Scores)

Many augmentation and co-construction users articulated an ethical commitment to preserving ownership of ideas. Students described internal boundaries that guided their use: *"If the ideas are mine and AI helps me express them better, I feel I'm still the author."* This theme directly supports the mediation pathway (AI-use → ownership → reasoning) and reinforces the HRF principle: humans retain responsibility for reasoning, even when AI assists in expression.

Conditional Co-Construction (Explains Mid-Range CRI Scores)

Co-construction users described working interactively with AI while maintaining critical agency. These students challenged AI suggestions, compared alternatives, or integrated multiple versions: *"It's like bouncing ideas around with something that responds, but I'm the one making the final choices."* This theme explains why co-construction users scored moderately on CRI, higher than replacement but lower than augmentation. AI-supported reasoning, but the greater reliance on iterative cycles diluted some of the profound reasoning benefits seen in augmentation.

Table 1: Thematic Coding Framework.

Participant Statement	Code	Primary Theme	Main Theme	Proposition
"It sounds good, but I can't justify it."	Lack of reasoning ownership	Reasoning Displacement	AI Replacement Weakens Cognition	Replacement AI reduces reasoning depth
"AI helps me find gaps in my argument."	Metacognitive prompting	AI-Enabled Reflection	Augmentation Strengthens Cognition	Augmentation improves reasoning
"If the ideas are mine, AI is just helping."	Ethical self-regulation	Ethical Augmentation	Ownership as Integrity	Cognitive ownership predicts integrity
"AI gives options; I choose what to keep."	Active human agency	Conditional Co-construction	Human–AI Agency Balance	Co-construction supports moderate reasoning
"I rely on AI too much sometimes."	Cognitive offloading	Reasoning Displacement	AI Overdependence	Overreliance undermines learning
"AI helped me reorganise my paragraph."	Structural support	AI-Enabled Reflection	Augmentation Benefits	Structural augmentation builds confidence

Integration: Explaining Convergence and Divergence

Table 2: Mixed Method Convergence Results.

Quantitative Finding	Qualitative Confirmation
Augmentation → highest reasoning	AI-enabled reflection strengthened cognitive processes.
Replacement → lowest reasoning	Reasoning displacement theme.
Augmentation → strongest ownership	Ethical augmentation practices.
AI-use → ownership → reasoning (mediation)	Students linked authorship with cognitive effort.

This convergence reinforces the validity of the HRF principle and underscores the centrality of ownership in reasoning-based academic integrity.

Table 3: Mixed Method Divergence Results.

Divergence	Quantitative Evidence	Qualitative Evidence	Interpretation & Implication
1. Overestimation of Ethical Confidence	Surveys showed **high self-reported ethical confidence** in AI use, suggesting students believed they were applying AI responsibly.	• Interviews revealed **uncertainty, inconsistent boundaries, and confusion** about what constitutes ethical AI use. • Students frequently said things like: *"I think it's okay… but I'm not sure where the line is."*	• Indicates that students **overestimate their ethical competence.** • Highlighting a gap between perceived and actual ethical understanding. • Strong need for **clear institutional guidelines**, training, and examples of acceptable AI use.
2. Perceived Learning vs. Actual Cognitive Engagement	Replacement users did not report significantly lower confidence or perceived learning in surveys.	• Students admitted feeling **more efficient but less cognitively engaged**, e.g., *"I get work done faster, but I don't think I'm actually learning."*	• Shows a disconnect between perceived productivity and **true reasoning development**. • Suggests risk of **cognitive offloading** not captured by quantitative measures. • Institutions must address this with **pedagogical interventions** that promote active reasoning.

These variances highlight areas where institutional guidance and pedagogical intervention are required. Integrated findings show that AI's effect on human reasoning depends not only on its integration but also on how it is engaged.

Augmentation highlights and reinforces human reasoning and sense of ownership.
Co-construction supports reasoning moderately.
Replacement undermines both reasoning and ownership.

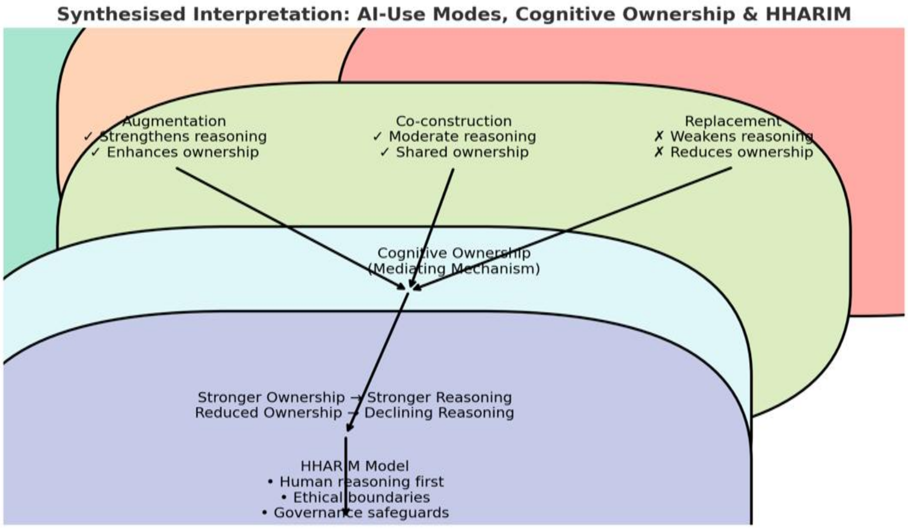

Figure 5: Synthesis Interpretation: AI-use Modes, Cognitive Ownership & HHARIM.

The SEM pathways and qualitative insights assemble on a core conclusion: cognitive ownership is the mechanism through which AI use influences human reasoning skills. In other words, when students feel that the work truly reflects their own skills and knowledge, their reasoning becomes stronger, more deliberate, and more ethically founded. When ownership diminishes, reasoning declines accordingly. These understandings directly formed the development of the Hybrid Human–AI Reasoning Integrity Model (HHARIM), which regards human reasoning as the essential foundation of academic writing and offers governance principles for HE.

7.0. Discussion

The findings directly support this inquiry by showing that the impact of AI on human reasoning depends not only on how it is used but also on the mode of engagement—Augmentation, Co-construction, or Replacement. This distinction clarifies an area where existing literature has mainly focused on plagiarism, textual misconduct, or general attitudes towards AI, while neglecting the cognitive effects of AI-supported writing processes (Selwyn, 2023; Cotton et al., 2023).

The literature review highlighted concerns about the displacement of cognitive effort when AI becomes a primary producer of content (Luckin, 2024; Williamson & Piattoeva, 2023), alongside emerging arguments that generative AI can enhance students' reflective capacity when used critically (Kasneci et al., 2023). The present findings provide strong empirical support for both positions but show that the outcome depends on how students interact with AI. Augmentation users demonstrated the highest reasoning scores and most substantial ownership—aligning with work suggesting that AI can scaffold metacognition and enhance learning skills when used as a supportive tool rather than an automated generator (Holmes et al., 2022). Co-construction, while beneficial, yielded moderate support for human reasoning, consistent with the literature identifying productive human–AI collaboration but also warning of over-reliance (Yang, 2022). Conversely, replacement users disclosed weaker reasoning and diminished cognitive ownership, reinforcing concerns about cognitive offloading and loss of authorship (Fischer et al., 2020; MacFarlane, 2019).

Significantly, this investigation expands existing research by empirically demonstrating, through structural equation modelling, that cognitive ownership is the key mechanism connecting AI-use mode to reasoning outcomes. This offers a more precise cognitive explanation than previous work, which tended to frame AI risks mainly in ethical or integrity terms. The qualitative findings provide nuance, showing that even HE students who believe they are using AI ethically may unknowingly diminish their cognitive involvement when relying on AI-generated text. This divergence highlights the overestimation of ethical competence, a phenomenon that is not thoroughly addressed in the current literature.

The preliminary results of this paper strongly reinforce the Human Reasoning First (HRF) principle proposed in the literature review: ethical academic writing must centre on preserving human reasoning as the primary intellectual engine. When HE students use AI as a reflective tool or structural scaffold, their reasoning remains active, and they retain the key generators of ideas. However, when AI replaces cognitive effort by generating arguments, explanations, or conceptual links, students lose control over their reasoning processes and experience lower levels of authorship. These findings sharpen the theoretical argument that reasoning integrity is the foundational component of academic integrity in the AI era. This position moves beyond traditional concerns with originality or plagiarism.

The relationship between human reasoning and ownership is not just correlational; it is sequential. The SEM findings and qualitative narratives confirm that greater ownership leads to stronger reasoning, while lesser ownership leads to weaker reasoning. This supports the idea that intellectual ownership is vital for deeper cognitive engagement (Bean, 2011). Therefore, the study confirms a reasoning-focused approach to academic integrity that emphasises human cognitive effort, not text production, as central to ethical AI use. The deviations identified in the mixed-methods analysis emphasise the challenges facing HE. However, students expressed high ethical confidence in surveys, yet interviews revealed uncertainty about the boundaries between ethical enhancement and unethical replacement. This gap suggests that students' internal ethical frameworks are inadequate for navigating the complexities of AI. As emphasised in the literature (Bretag, 2019; Eaton, 2023), academic integrity in modern contexts requires more than mere compliance; it needs explicit pedagogical guidance on human reasoning processes, authorship boundaries, and ethical judgment.

The preliminary findings suggest three key implications:
Institutions must provide more straightforward, reasoning-focused AI guidelines.
Generic warnings about plagiarism are insufficient. The students require explicit demonstrations of the difference between thinking with AI and thinking via AI.
Pedagogy should highlight cognitive ownership. Learning activities must require students to articulate their human reasoning steps, justify their decisions, and reflect on their thinking processes, practices that discourage the use of replacement-based AI.
AI literacy training must integrate ethical reasoning. Students need structured opportunities to critique AI outputs, identify illusions, and evaluate the epistemic reliability of generative AI models.
These implications are consistent with emerging calls in the literature for "critical AI literacy" and for academic institutions to develop governance frameworks tailored to student reasoning processes (Holmes & Tuomi, 2022).

8.0. Conclusion

This analysis explored how different modes of generative AI use—augmentation, co-construction, and replacement shape students' reasoning skills and cognitive ownership in higher education. Across both quantitative and qualitative approaches, a clear pattern emerged: AI's influence on learning is not determined by its existence but by how students engage with it. Augmentation enhanced human reasoning by supporting metacognitive reflection, co-construction offered moderate cognitive benefits, and replacement undermined both reasoning and perceived ownership. These findings provide strong empirical support for redefining academic integrity as human reasoning integrity, prioritising human-led cognition and intellectual autonomy over text originality alone.

A key theoretical contribution of this study is the identification of cognitive ownership as the central mechanism relating AI-use mode to reasoning outcomes. Structural equation modelling demonstrated a sequential pathway AI use influences ownership, ownership influences reasoning, and reasoning shapes perceptions of integrity. This clarifies why augmentation improves learning while replacement hinders it, filling a critical gap in the emerging literature on human–AI interaction.

The findings informed the development of the Hybrid Human–AI Reasoning Integrity Model (HHARIM), which offers a pedagogical and governance-focused framework for ethical AI integration. HHARIM positions human reasoning as the foundation of academic practices, sets boundaries for AI use, and emphasises the importance of human-reasoning-centred policies and AI literacy programmes. As most higher education institutions navigate the rapid normalisation of generative AI, this framework provides a practical and theoretically grounded guide to maintaining intellectual agency. However, this research also has limitations, including reliance on self-reported data, a cross-sectional design, and a sample drawn from UK institutions. Future research should employ longitudinal, experimental, and cross-cultural methods to examine how AI use patterns evolve over time and how they impact cognitive processes. Nonetheless, the findings highlight an important message: safeguarding human reasoning not merely detecting plagiarism is vital for fostering meaningful learning in the age of generative AI.

References

Andrews, R. (2010) *Argumentation in higher education: Improving practice through theory and research*. Routledge.

Bandura, A. (2001) 'Social cognitive theory: An agentic perspective', *Annual Review of Psychology*, 52(1), pp. 1–26. Available at: https://doi.org/10.1146/annurev.psych.52.1.1

Bean, J.C. (2011) *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom*. 2nd edn. Jossey-Bass.

Braun, V. and Clarke, V. (2021). 'One size fits all? What counts as quality practice in (reflexive) thematic analysis', *Qualitative Research in Psychology*, 18(3), pp. 328–352. Available at: https://doi.org/10.1080/14780887.2020.1769238

Bretag, T. (2019) 'Academic integrity: Navigating the "new normal"', *International Journal for Educational Integrity*, 15(1), Article 1. Available at: https://doi.org/10.1007/s40979-019-0040-9

Crompton, H. and Burke, D. (2023) 'Artificial intelligence in higher education: The state of the field', *International Journal of Educational Technology in Higher Education*, 20(1), Article 22. Available at: https://doi.org/10.1186/s41239-023-00392-8

Cotton, D.R.E., Cotton, P.A. and Shipway, J.R. (2023) 'Chatting and cheating: Ensuring academic integrity in the era of ChatGPT', *Innovations in Education and Teaching International*, 61(2), pp. 228–239. Available at: https://doi.org/10.1080/14703297.2023.2190148

Creswell, J.W. and Plano Clark, V.L. (2018) *Designing and conducting mixed methods research*. 3rd edn. SAGE.

Dawson, P. (2020) 'Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education', *Higher Education Research & Development*, 39(1), pp. 44–58. Available at: https://doi.org/10.1080/07294360.2019.1680950

Dwivedi, Y.K. et al. (2024) 'The future of generative AI: Academic perspectives on opportunities and challenges', *International Journal of Information Management*, 74, Article 102694. Available at: https://doi.org/10.1016/j.ijinfomgt.2023.102694

Eaton, S.E. (2023) *Ethical AI and academic integrity in higher education*. University of Calgary Press.

Fischer, K., Boone, G. and Neubert, J. (2020) 'Cognitive offloading and its implications for learning with digital technologies', *Educational Psychology Review*, 32, pp. 1–21. Available at: https://doi.org/10.1007/s10648-019-09500-5

Harper, R., Bretag, T. and Rundle, K. (2021) 'Detecting contract cheating: Examining the role of assessment type', *Higher Education Research & Development*, 40(2), pp. 265–279. Available at: https://doi.org/10.1080/07294360.2020.1724899

Holmes, W., Bialik, M. and Fadel, C. (2022) *Artificial intelligence in education: Promises and implications for teaching and learning*. 2nd edn. Centre for Curriculum Redesign.

Holmes, W. and Tuomi, I. (2022) 'Critical AI literacy in higher education: A framework for teaching and learning', *AI & Society*, 37, pp. 1207–1221. Available at: https://doi.org/10.1007/s00146-021-01270-5

Kasneci, E. et al. (2023) 'ChatGPT for education and research: Opportunities, threats, and strategies', *Frontiers in Education*, 8, Article 1197990. Available at: https://doi.org/10.3389/feduc.2023.1197990

Kline, R.B. (2016) *Principles and practice of structural equation modelling*. 4th edn. Guilford Press.

Lea, M.R. and Street, B. (1998) 'Student writing in higher education: An academic literacies approach', *Studies in Higher Education*, 23(2), pp. 157–172. Available at: https://doi.org/10.1080/03075079812331380364

Lee, D. et al. (2024) 'The impact of generative AI on higher education learning and teaching: A study of educators' perspectives', *Computers and Education: Artificial Intelligence*, 6, Article 100221. Available at: https://doi.org/10.1016/j.caeai.2024.100221

Luckin, R. (2024) *AI for learning: Why artificial intelligence matters for education*. Routledge.

Macfarlane, B., Zhang, J. and Pun, A. (2014) 'Academic integrity: A review of the literature', *Studies in Higher Education*, 39(2), pp. 339–358. Available at: https://doi.org/10.1080/03075079.2012.709495

Morgan, D.L. (2007) 'Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods', *Journal of Mixed Methods Research*, 1(1), pp. 48–76. Available at: https://doi.org/10.1177/2345678906292462

Popenici, S. and Kerr, S. (2017) ‘Exploring the impact of artificial intelligence on teaching and learning in higher education’, *Research and Practice in Technology Enhanced Learning*, 12(1), Article 22. Available at: https://doi.org/10.1186/s41039-017-0062-8

Searle, A., Gašević, D. and Dawson, S. (2024) ‘Understanding student use of generative AI: A learning analytics perspective’, *Computers & Education*, 204, Article 104893. Available at: https://doi.org/10.1016/j.compedu.2023.104893

Selwyn, N. (2023) *Should robots replace teachers? AI and the future of education*. 2nd edn. Polity Press.

Shute, V.J. et al. (2021) ‘The role of AI in supporting metacognition and learning’, *Educational Psychologist*, 56(4), pp. 250–262. Available at: https://doi.org/10.1080/00461520.2021.1956566

Sparrow, B., Liu, J. and Wegner, D.M. (2011) ‘Google’s effects on memory: Cognitive consequences of having information at our fingertips’, *Science*, 333(6043), pp. 776–778. Available at: https://doi.org/10.1126/science.1207745

Thompson, R. and Lee, M.J. (2012) ‘Talking with students through screencasting: Experimentations with video feedback to improve student learning’, *Journal of Interactive Technology and Pedagogy*, 1(1), pp. 1–16. Available at: https://jitp.commons.gc.cuny.edu/talking-with-students-through-screencasting-experimentations-with-video-feedback-to-improve-student-learning/

UNESCO (2023) *Guidance for generative AI in education and research*. UNESCO Publishing.

Weller, M. (2020) *25 years of ed tech*. Athabasca University Press.

Williamson, B. and Eynon, R. (2020) ‘Automating education: AI technologies and the governance of learning’, *Learning, Media & Technology*, 45(1), pp. 1–6. Available at: https://doi.org/10.1080/17439884.2019.1667823

Wingate, U. (2012) ‘Using academic literacies and genre-based models for academic writing instruction: A “literacy” journey’, *Journal of English for Academic Purposes*, 11(1), pp. 26–37. Available at: https://doi.org/10.1016/j.jeap.2011.11.006

Yang, S. (2022) ‘Human–AI collaboration in writing: Opportunities and cognitive risks’, *Computers & Education*, 187, Article 104574. Available at: https://doi.org/10.1016/j.compedu.2022.104574

Zimmerman, B.J. (2002) ‘Becoming a self-regulated learner: An overview’, *Theory into Practice*, 41(2), pp. 64–70. Available at: https://doi.org/10.1207/s15430421tip4102_2

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 45-54
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1770

Influence Of Personal Branding on Entrepreneurial Success of Fitness Coaches in the UK.

[1]Michelle Pazmino- International SOS
[2]Olugbenga Adeyanju Akintola; [3]Emmanuel Nwachukwu; [4]Mosunmola Adeyeye
[2, 3, & 4]Northumbria University, London

Abstract
Personal branding allows entrepreneurs to develop strong relationships with their customers and drive emotional affinity, trust, and loyalty. Nonetheless, the specific strategies that lead to entrepreneurial success in the fitness industry remain less clear. The present study evaluates the influence of three personal branding strategies (authenticity, attractiveness, and credibility) on the tribalist entrepreneurial success of fitness coaches in the UK. Drawing on data from 169 surveys, the study reveals that when examining each dimension separately, authenticity has a significant influence ($\beta = 0.612$, $p < .001$) on entrepreneurial success, while credibility has a suggestive influence although below the significance threshold ($\beta = 0.186$, $p = .068$), and attractiveness has no significant influence on entrepreneurial success ($\beta = 0.006$, $p = .944$). Moreover, the composite variable of personal branding strategies indicated a significant influence on entrepreneurial success ($\beta = 0.796$, $p < .001$). These findings allow practical recommendations to be provided to fitness coaches to develop an integrated personal branding strategy that encompasses the three dimensions evaluated in this research to maximise their entrepreneurial outcomes. Additional research utilising qualitative or mixed methodology and focusing on both customers' and fitness coaches' perspectives would be valuable to obtain a comprehensive understanding of the influence of personal branding strategies on the tribalist entrepreneurial success of fitness coaches in the UK.

Keywords: Personal branding, Entrepreneurial success, Fitness coaches, Authenticity, Credibility, Attractiveness.
Wordcount: 216

1.0 Introduction

The rise of solo service entrepreneurship has fundamentally changed how individuals build and sustain businesses. Across the UK, there has been an increase in the number of people who now work for themselves, offering services from fitness coaching to consulting and therapy (ONS, 2024). Unlike traditional businesses that can differentiate through product features or brand reputation, expertise, and ability to connect with customers directly determine whether they succeed or fail (Resnick et al., 2016). This challenge intensifies in commoditised service markets where services appear functionally similar, and customers struggle to assess quality before purchase. Personal branding then becomes the strategic presentation of one's identity, expertise, and values to differentiate from competitors (Shepherd, 2005). Entrepreneurs can remove uncertainty by deliberately crafting and communicating their personal brand to build trust (Gorbatov et al., 2019). However, three critical gaps limit our understanding. First, existing studies examine branding strategies in isolation rather than comparing their relative effects. In addition, most research reviewed focuses on entrepreneurs' perspectives rather than customers' perceptions, despite brand success ultimately depending on customer evaluation. Third, evidence conflicts on whether strategies operate independently or must be combined for effectiveness.

The fitness coaching industry provides a suitable context for addressing these gaps. The UK fitness market reached £5.9billion in 2023, with approximately 183,300 people working in sports and fitness occupations (Mintel, 2024; Statista, 2023). The COVID-19 pandemic forced coaches to develop online offerings, yet many struggle to build client relationships remotely and differentiate themselves in saturated digital spaces (Kulkarni, 2024; North et al., 2020). Beyond its practical relevance, this context suits our theoretical inquiry for three reasons. (i) High commoditisation and information asymmetry create conditions where signalling becomes crucial for customer decision-making (Connelly et al., 2011). Customers cannot easily assess coaching quality before purchase, making them reliant on observable brand signals. (ii) Severe capacity constraints mean coaches cannot scale through volumes, so success must derive from relationship quality rather than quantity. (iii) The visibility of coaches' physical bodies provides natural variation in personal presentation strategies, enabling empirical comparison of authenticity, credibility, and attractiveness effects.

Research identifies authenticity, credibility, and attractiveness as foundational personal branding strategies (Scheidt & Henseler, 2018; Thompson-Whiteside et al., 2018; Hutson, 2013), yet evidence on their respective importance remains mixed. Some studies emphasise authenticity as foundational (Kristiansen & Williams, 2015; Vitelar, 2019), others highlight credibility (Wei et al., 2022), whilst still others focus on attractiveness (Hutson, 2013). Moreover, most research examines athletes or corporate professionals rather than solo entrepreneurs operating capacity-constrained ventures. Critically, studies typically adopt the entrepreneur's perspective, leaving customers' views largely unexplored despite customers ultimately determining brand success through their perception and choice. This study addresses these gaps by examining how authenticity, credibility, and attractiveness strategies influence entrepreneurial success from customers' perspectives in the context of UK fitness coaches.

2.0. Literature Review

Solo service entrepreneurs operate in markets characterised by pronounced information asymmetry, where customers cannot directly observe service quality, expertise or commitment before purchase (Mauri et al., 2018). This uncertainty proves particularly acute in fitness coaching, where clients must commit substantial time and financial resources based on limited information about whether a coach will deliver promised results. Personal branding emerges as entrepreneurs' strategic response to this challenge, representing deliberate efforts to construct and communicate a distinctive professional identity that signals quality whilst fostering emotional connections with potential customers. Two complementary theoretical perspectives explain how these strategies translate into entrepreneurial outcomes, addressing cognitive and affective pathways that operate simultaneously to shape customer perceptions and behaviours.

Signalling theory, which originates from an analysis of labour market dynamics, provides the foundation for understanding how personal branding reduces customer uncertainty (Spence, 1973). The theory proposes that informed parties communicate unobservable qualities through observable signals, with signal credibility depending on cost, which makes imitation by low-quality actors unprofitable (Connelly et al., 2011). This cost structure creates separating equilibria wherein high-quality actors find signalling worthwhile while low-quality actors do not, enabling customers to distinguish between providers based on observable behaviours rather than unverifiable claims. Mauri et al. (2018) analysed 3,847 Airbnb listings and found that host reputation signal reduced booking uncertainty by 34% in contexts where quality becomes apparent only through sustained interaction.

Personal branding strategies function as quality signals through distinct mechanisms that differ in cost and the specific qualities they communicate. Authenticity signals trustworthiness through costly vulnerability, as entrepreneurs who consistently share genuine experiences, including struggles and flaws, engage in behaviour that untrustworthy actors would rationally avoid given the reputational risk of exposure (Thompson-Whiteside et al., 2018). The mechanism operates through what the signalling theory terms as a separating equilibrium. Entrepreneurs who share personal failures and setbacks demonstrate confidence in their character, effectively communicating that they have nothing to hide. Dishonest entrepreneurs would avoid such disclosure because the exposure of inconsistency between their projected image and actual character would damage their reputation and undermine their business. The willingness to be vulnerable credibly communicates honesty, not because the content itself proves trustworthiness, but because sharing carries asymmetric risk depending on the sender's actual character. Campagna et al.'s (2022) meta-analysis of 47 studies involving 12,483 participants found perceived authenticity strongly predicted trust ($r=.68$, $p<.001$) and customer loyalty ($r=.61$, $p<.001$), whilst Thompson-Whiteside et al., (2018) interview revealed that 92% of female entrepreneurs reported authentic posts outperformed success-only narratives despite 67% expressing vulnerability concerns. This paradox, where authenticity works precisely because it feels risky, validates signalling theory's core logic that costliness creates credibility.

Credibility signals competence through demonstrated expertise and verifiable qualifications, operating through investment cost rather than vulnerability risk. Educational credentials, professional certifications, and demonstrated knowledge represent costly-to-acquire indicators that are difficult to fake, making them reliable proxies for actual competence. A study of 412 social media influencers found credibility indicators increased follower trust by 41% and purchase intentions by 28%, confirming that credentials function as a reliable signal in contexts where technical expertise is salient to customer decisions (Wei et al., 2022).

However, evidence on credibility's influence is less consistent than authenticity findings, particularly in relationship-intensive service contexts. Qualitative research found that entrepreneurs who led with credentials were sometimes perceived as less relatable, creating professional distance rather than personal connection with potential customers (Thompson-Whiteside et al., 2018). This suggests a potential trade-off between competence signalling and relationship building. In services such as fitness coaching, where customers seek both expertise and personal connection, credentials may establish baseline legitimacy but contribute less to the emotional bonds that characterise loyalty. Customers may discount expertise signals when they prioritise shared values and personal rapport over technical knowledge, particularly given that baseline competence is often assumed once minimum qualifications are met. This pattern suggests credibility operates primarily through cognitive pathways, effectively addressing whether someone can deliver results but offering limited capacity for the emotional bonding that characterises tribalist entrepreneurial success.

Attractiveness signals professionalism through visual presentation and, in fitness contexts, physical fitness demonstrates commitment and expertise. Achieving and maintaining an athletic physique requires sustained effort, discipline and knowledge that uncommitted actors would find prohibitively costly to fake. Customers, therefore, interpret physical attractiveness as evidence that the coach practices what they preach and possess the discipline necessary to guide others toward similar outcomes. An ethnographic study involving 26 trainers and 25 clients found that 89% of clients initially selected trainers based on physical appearance, viewing athletic physiques as tangible proof of competence (Hutson, 2013). The economic implications were substantial, with trainers in the top appearance quartile charging 47% higher rates than bottom quartile trainers, even with strong signalling in experience and credentials. Experimental research with 324 participants further demonstrated that strong visual brand signals enable 23% price premiums, suggesting attractiveness carries real market value (Thai & Wang, 2020).

The relationship between visual presentation and underlying service quality is less direct than other signals. Attractiveness presents both theoretical and empirical ambiguity that distinguishes it from authenticity and credibility maps onto competence. Attractiveness may signal general professionalism without conveying specific information about coaching ability or interpersonal skills. This is supported by a survey carried out on 389 sports fans, the research findings showed no significant direct effect of perceived attractiveness on brand connection ($\beta=.08$, $p=.23$) or brand love ($\beta=.11$, $p=.14$), despite significant bivariate correlations (Zhou et al., 2020). Regression analysis revealed attractiveness operated primarily through perceived authenticity (indirect effect: $\beta=.28$, $p<.001$) rather than exerting independent influence. This suggests attractiveness may function differently than signalling theory initially predicts.

Furthermore, three competing explanations emerge (i) attractiveness may operate as a hygiene factor, meeting baseline expectations necessary for consideration but insufficient for differentiation once thresholds are met. (ii) Attractiveness may function as an amplifier of other strategies rather than an independent signal, enhancing authentic messages without generating connection on its own. (iii) Attractiveness effects may be stage-dependent, which would matter more for initial attention but less for relationship development, which would explain why ethnographic research found selection effects (Hutson, 2013) while survey research found no relationship effects (Zhou et al., 2020). These explanations require empirical testing.

Self-expansion theory posits that individuals possess intrinsic motivation to enhance their sense of self by incorporating new resources, experiences and identities from close relationships, which fosters loyalty and complements signalling theory by reducing uncertainty (Aron & Aron, 1986; Aron et al., 2013). Extended to brand contexts, consumers establish relationships with brands that encourage personal growth, incorporating brand attributes into their own identities (Kerviler & Rodriguez, 2019). When fitness coaches' personal brands embody aspirational attributes, customers experience self-growth through association, seeing themselves reflected in authentic narratives or aspiring to embody demonstrated qualities. A survey of 487 fitness consumers found perceived authenticity predicted parasocial relationships ($\beta=.71$, $p<.001$), which in turn predicted loyalty ($\beta=.68$, $p<.001$), confirming the mechanism operates in fitness contexts where customers actively pursue identity transformation (Li et al., 2023). Authentic brands provide clear, consistent identities for customer incorporation into self-concept, transforming clients into community members whose brand relationships enhance their sense of who they are becoming rather than simply providing a service transaction. Whilst evidence links each strategy to positive outcomes, no study has examined all three simultaneously from customers' perspectives to determine their relative contributions.

This theoretical framework suggests strategies should differ in their effects given distinct pathway activations, yet this remains untested. And while the dual-pathway model implies collective deployment may enhance influence beyond individual contributions, this proposition lacks empirical support in solo entrepreneurship contexts. Hence, the research questions below;

RQ1: What relationships exist between personal branding strategies and tribalist entrepreneurial success?
RQ2: To what extent do personal branding strategies differ in their individual effects on tribalist entrepreneurial success?
RQ3: Do combined personal branding strategies influence tribalist entrepreneurial success beyond individual strategy effects?

The three strategies function as signals that reduce customer uncertainty, suggesting each should positively influence tribalist entrepreneurial success. However, their effects should differ in magnitude. Authenticity activates both cognitive and affective pathways, reducing uncertainty while simultaneously providing identity resources for self-expansion. Credibility operates primarily through cognitive pathways, establishing competence but offering weaker material for emotional connection. Attractiveness may function as a hygiene factor or amplifier rather than an independent driver of relationship-based success. Furthermore, deploying multiple strategies should activate complementary pathways, generating collective influence beyond what individual strategies contribute separately.

H1a: Authenticity positively influences tribalist entrepreneurial success.
H1b: Credibility positively influences tribalist entrepreneurial success.
H1c: Attractiveness positively influences tribalist entrepreneurial success.
H2: Personal branding dimensions differ in their relative influence on tribalist entrepreneurial success.
H3: Combined personal branding strategies positively influence tribalist entrepreneurial success.

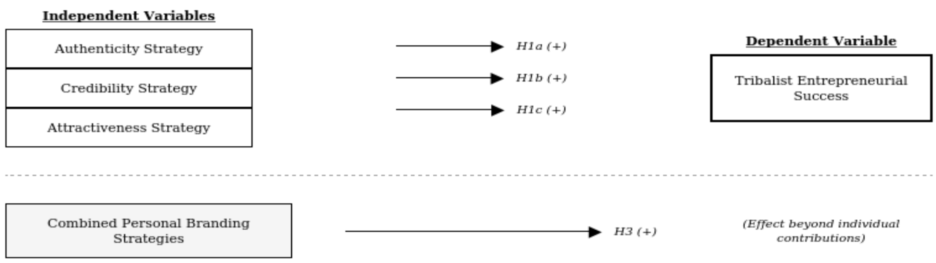

Figure 1. Conceptual Framework

3.0. Method

3.1. Research Design and Sample

The research utilised a correlational design, as the primary focus was to assess the influence of the independent variables (personal branding strategies-Authenticity, Attractiveness, and Credibility) on the dependent variable (Entrepreneurial Success) without altering any of the variables involved. The target population comprised UK residents who actively engage with fitness coaching brands on social media and online platforms. Because these individuals are dispersed across platforms with no centralised register, snowball sampling was employed, consistent with similar social media brand research (Li et al., 2023; Thompson-Whiteside et al., 2018). The survey was distributed through fitness-focused social media groups on Instagram, Facebook, and LinkedIn. Participants were encouraged to share the link with others who met the eligibility criteria. To be included, participants had to be aged 18 or above, reside in the UK, and actively follow or engage with at least one fitness coach on social media. The final sample comprised 169 respondents, exceeding the 107 minimum recommended for multiple regression with three predictors at 80% power and medium effect size (Cohen, 1992).

3.2. Instruments

Structured questionnaire was employed to collect data for the study. The Questionnaire measured the study variables using 7-point Likert scales (1 = Strongly Disagree to 7 = Strongly Agree). The Fitness Coaches Personal Branding Scale (FC-PBS $\alpha = 0.78$) comprised 13 items measuring three dimensions. Original scale items were reworded to reflect fitness coaching contexts while retaining the underlying construct definitions and response format. The attractiveness subscale (3 items) drew on McCroskey and McCain (1974, as cited in Sakib et al., 2020) and Ohanian (1990, as cited in Zhou et al., 2020), measuring visual brand appeal (e.g.,

"I am more likely to engage with a fitness coach's brand if it is designed well"). The authenticity subscale (5 items), adapted from Lee and Eastin (2021), assessed perceived genuineness (e.g., "I am more likely to engage with a fitness coach's brand if it comes off as very genuine"). The credibility subscale (5 items) combined source credibility dimensions from McCroskey and Teven (1999, as cited in Sakib et al., 2020) and Ohanian (1990, as cited in Zhou et al., 2020) with information quality indicators from Kim (2022) (e.g., "I am more likely to engage with a fitness coach's brand if it provides reliable fitness information").
Five items adapted from Pontinha and Coelho do Vale (2020) measured tribalist entrepreneurial success, defined as the degree to which customers perceive emotional affinity and community connection with a fitness coach's brand (Angel et al., 2018). Items were reworded to reference fitness coaches specifically (e.g., "I feel an emotional connection to the fitness coach").

3.3. Data Collection and Analyses

The questionnaires were distributed online, hosted on Google Forms. The survey link was distributed through fitness-focused social media groups on Instagram, Facebook, and LinkedIn, and participants were encouraged to share the link with others who met the eligibility criteria. The survey opened with an introduction explaining the research purpose, followed by demographic questions and scale items organised by construct. Prior to participation, respondents received a debrief sheet outlining the study's purpose and data handling practices. Informed consent was obtained, with participants acknowledging the voluntary nature of their involvement and their right to withdraw without consequence. In accordance with GDPR requirements, no personally identifiable information was collected, and all responses were stored securely with access restricted to the research team. No incentives were offered.

The analyses were conducted using IBM SPSS Statistics at $\alpha = 0.05$. Pearson product-moment correlations tested bivariate relationships between each personal branding strategy and entrepreneurial success (H1). Multiple linear regression assessed each strategy's unique contribution while controlling for the other two (H2). Simple linear regression using a composite personal branding variable tested collective influence (H3).

4.0. Results and Discussion

4.1. Results

H1a, H1b, and H1c predicted that authenticity, credibility, and attractiveness would each positively influence tribalist entrepreneurial success.

The results in Table 1 suggested strong relationships between personal branding strategies (attractiveness, authenticity, and credibility) and the entrepreneurial success of fitness coaches in the UK. H1a was supported: authenticity positively correlated with entrepreneurial success ($r = .645$, $p < .001$), representing a large effect. H1b was supported: credibility positively correlated with entrepreneurial success ($r = .547$, $p < .001$), representing a large effect. H1c was supported: attractiveness positively correlated with entrepreneurial success ($r = .416$, $p < .001$), representing a medium effect (Cohen, 1988).

Table 1. Correlation Matrix on Personal Branding Strategies and Entrepreneurial Success of Fitness

	Attractiveness Strategy	Authenticity Strategy	Credibility Strategy	Entrepreneurial Success
Attractiveness Strategy	1			
Authenticity Strategy	.631**	1		
Credibility Strategy	.496**	.734**	1	
Entrepreneurial Success	.416**	.645**	.547**	1

**. Correlation is significant at the 0.01 level (2-tailed).

H2 predicted that personal branding dimensions differ in their relative influence on entrepreneurial success.
Table 2 showed a strong positive relationship between the independent variables (attractiveness, authenticity, and credibility) and the entrepreneurial success of fitness coaches in the UK ($R = .654$; $R^2 = .427$; Adj $R^2 = .417$; $F(3, 165) = 41.067$; $p < .001$). This showed that 42.7% of the variance in entrepreneurial success can be attributed to these personal branding strategies, representing a large effect size (Cohen, 1988). In addition, Table 2 denoted that the model is statistically significant and therefore at least one of the personal branding strategies is a significant predictor of entrepreneurial success of fitness coaches in the UK. Moreover, the results indicated that authenticity is the most significant predictor of entrepreneurial success ($\beta = 0.612$, $p < .001$), demonstrating that an authentic brand will most likely lead to entrepreneurial success. The credibility strategy ($\beta = 0.186$, $p = .068$) is slightly above the threshold for significance.

Therefore, the credibility of the brand could modestly impact the entrepreneurial success of the fitness coach; however, this requires further research to confirm the influence. Lastly, the attractiveness strategy ($\beta = 0.006$, $p = .944$), as defined in this research, does not have a significant influence on the entrepreneurial success of fitness coaches in the UK. H2 is supported. The dimensions differ substantially in their relative influence on entrepreneurial success. Authenticity demonstrated a strong and significant influence, credibility showed a modest but non-significant influence, and attractiveness showed no significant influence. These findings indicate that authenticity is a pivotal element when building a personal brand to enhance the entrepreneurial success of fitness coaches.

Table 2. Model Summary of the Multiple Regression Analysis for the Individual Contribution of Personal Branding Strategies to the Entrepreneurial Success of Fitness Coaches in the UK.

Predictor	β	p
Authenticity	0.612	<.001
Credibility	0.186	.068
Attractiveness	0.006	.944

Note. $R = .654$, $R^2 = .427$, Adjusted $R^2 = .417$, $F(3, 165) = 41.067$, $p < .001$. Dependent variable:

Entrepreneurial success

H3 predicted that combined personal branding strategies positively influence tribalist entrepreneurial success.

The results in Table 3 illustrated that the composite variable of the three personal branding strategies (attractiveness, authenticity, and credibility) is a significant predictor of entrepreneurial success of fitness coaches in the UK ($R=.615$; $R2=.375$; Adj $R2=.375$; $F(1, 167) =101.593$; $\beta = 0.796$; $p <.05$). This suggested that 37.8% of the variance of entrepreneurial success can be explained by personal branding strategies collectively, representing a large effect size (Cohen, 1988). H3 is supported; the strategies attractiveness, authenticity, and credibility are collectively of significant importance when building a personal brand to drive entrepreneurial success of fitness coaches in the UK.

Table 3. Model Summary of the Multiple Regression Analysis for the Combined Contribution of Personal Branding Strategies to the Entrepreneurial Success of Fitness Coaches in the UK.

Predictor	β	p
Personal branding	0.796	<.001

Note. $R = .615$, $R^2 = .378$, Adjusted $R^2 = .375$, $F(1, 167) = 101.593$, $p < .001$. Dependent variable is Entrepreneurial success.

4.2. Discussion

Examining each dimension, authenticity was found to have a positive and significant influence on entrepreneurial success. This aligns with the assertion of Scheidt and Henseler (2018), that authenticity is key when building a personal brand, as it influences customers' perceptions and judgements. Moreover, Lunardo et al. (2015) explain that authentic brands that prioritise sincerity and trustworthiness are more likely to establish emotional affinity with their customers. Similarly, Kowalczyk and Pounders' (2016) findings suggest that customers seek personal brands that share authentic posts and insights into their lives rather than brands that focus mainly on their careers and promotional content. This was evidenced in the study through representative items such as *"I am more likely to engage with a fitness coach's brand if it not only posts the good in their life but also about hardships" and "I am more likely to engage with a fitness coach's brand if it talks about their flaws and is not ashamed for showing them to the public".*

On the other hand, Scheidt et al. (2020) and Thompson-Whiteside et al. (2018) affirm that it can be challenging to express one's authenticity as society puts pressure to adhere to the norm and satisfy the market. Moreover, Gehl (2011) expresses that individuals aiming to build their personal brands are forced to expose their personal lives in an effort to demonstrate authenticity. In the study conducted by Thompson-Whiteside et al. (2018), respondents expressed their fear of rejection when trying to be authentic, as their audiences may not like them and judge them. Nevertheless, Thompson-Whiteside et al., (2018) highlighted the importance of developing resilience and networking skills as coping strategies, and emphasised that perfection and being liked by everyone is not realistic. Authenticity is necessary for developing connection as it reflects all of the individual traits in one dimension, resulting in perceived emotional affinity and customer loyalty (Kucharska et al., 2020; Kowalczyk & Pounders, 2016). This suggests that fitness coaches in the UK will be more likely to experience greater levels of tribalist entrepreneurial success by integrating authenticity into their personal brands (Kucharska et al., 2020).

The results, however, revealed a suggestive yet non-significant relationship between credibility and entrepreneurial success. Lunardo et al. (2015) emphasise that customers tend to develop a stronger connection with brands which are perceived as credible and capable of meeting their needs and goals. Thompson-Whiteside et al. (2018) also identified the need for maintaining professionalism in the personal brand to be seen as credible. This was evidenced through key representative items including *"I am more likely to engage with a fitness coach's brand if it looks knowledgeable", "I am more likely to engage with a fitness coach's brand if it provides useful information on fitness", and "I am more likely to engage with a fitness coach's brand if it provides reliable information on fitness".* The respondents in the study indicated a high level of agreement with the mentioned constructs, with mean scores of 6.20, 6.17, and 6.25, respectively, on a 7-point Likert scale, suggesting that customers also consider the credibility of a personal brand; however, it may not be of relative importance for establishing a connection with the coach. The report of Mauri et al. (2018) underpins these findings that the credibility of a brand is important in bridging the trust gap in the transaction. Likewise, personal brands with a great reputation will always have higher popularity (Mauri et al., 2018). This study denoted that credibility by itself may not have a significant influence on entrepreneurial success; however, credibility of a personal brand can help enhance the trustworthiness of the fitness coach (Campagna et al., 2022), which is considered a main dimension of authenticity, contributing then to higher degrees of entrepreneurial success. Moreover, trustworthiness also influences how credible a personal brand is perceived (Phung & Qin, 2018), indicating that credibility and authenticity are closely intertwined.

The results of the research also showed that attractiveness does not influence the entrepreneurial success of fitness coaches. Comparably, Zhou et al. (2020) found no significant effect of attractiveness on the consumers' connection and trust for the brand. Zhou et al. (2020) explain that it may be challenging to judge the attractiveness of a brand as it is mostly unlikely for an athlete to choose an unattractive personal brand design as they will try to meet their customers' demands in terms of symbolic and visual appeal. Furthermore, Hutson (2013) manifests that the physical appearance of fitness coaches is what gives them credibility in their work and allows them to be successful. The participants of his study agreed that the appearance, attractiveness or physical ability of the coach is perceived as a commodity that represents their commitment to being physically fit and healthy (Hutson, 2013) and therefore trustworthy trainers for their clients. Sakib et al. (2020) observed that consumers are more engaged with a brand when perceived as more credible and attractive and that their credibility may be naturally caused by their attractiveness. This suggests that although consumers may seek to engage with attractive brands, it is not of most importance for tribalist entrepreneurial success on its own.

In summary, the findings of this study suggest that, relatively, authenticity is the only strategy that significantly and positively influences entrepreneurial success, while credibility and attractiveness on their own are not enough to influence the entrepreneurial success of fitness coaches in the UK. Nonetheless, collectively, attractiveness, authenticity, and credibility positively and significantly influence tribalist entrepreneurial success of fitness coaches in the UK. Gorbatov et al. (2019) emphasise that there is ample proof to suggest that personal branding leads to increased visibility, credibility, and reputation, among other benefits. Carlson and Donavan (2013) identified that in the field of sports, the personal brand of an athlete positively influences the degree to which a customer develops an emotional connection with the athlete. In addition, Reimann et al. (2012) explain that strong brand relationships can be accounted for by self-expansion, where individuals seek to grow and expand by associating with a brand that reflects their values and aspirations. Fitness coaches portraying authentic, credible and appealing content through their personal brands can offer an opportunity to enhance their customers' sense of self (Zhou et al., 2020). Consequently, as the brand resonates with its audience and customers, it will encourage emotional connection with the brand and brand loyalty (Zhou et al., 2020). As a result, fitness coaches in the UK must leverage an integrated personal branding strategy, including authenticity, credibility and attractiveness, on social media to strengthen their relationship with their customers (Chen, 2013) and achieve greater tribalist entrepreneurial success.

5.0. Conclusion and Recommendation

This research addressed the relative and collective influence of three personal branding strategies: authenticity, attractiveness, and credibility, on the entrepreneurial success of fitness coaches in the UK. Prior research has suggested that personal branding is essential in building strong relationships with customers, customer retention and loyalty (Dašić et al., 2021; Kristiansen & Williams, 2015; Mousavi-Jad et al., 2021; Staškevičiūtė-Butienė et al., 2014; Thompson-Whiteside et al., 2018). The results of this study revealed that, relatively, only authenticity showed a positive and significant influence, while credibility had a suggestive influence, and attractiveness had no significant influence on entrepreneurial success. On the other hand, authenticity, attractiveness, and credibility collectively have a positive and significant influence on the tribalist entrepreneurial success of fitness coaches in the UK. Hence, the importance of strategically building an integrated personal brand to stand out in a competitive market and enhance the entrepreneurial success of fitness coaches in the UK.

Based on these findings, these recommendations are made, fitness coaches in the UK should focus on authenticity, making sure they present themselves as a transparent and genuine brand, sharing original content which represents not only the good in their lives but also the hardships, and not being ashamed of sharing their flaws with the public. This will help fitness coaches to be perceived as trustworthy and authentic, therefore, strengthening the relationship with their customers. In addition, fitness coaches can enhance their perceived trustworthiness by being knowledgeable and providing useful and reliable information on fitness. As well as making their customers feel comfortable and utilising a visually appealing brand design. Credibility and attractiveness are positively correlated with authenticity; therefore, they can further enhance the connection and trust customers have towards the brand. By adopting all three strategies collectively, fitness coaches in the UK can position themselves as industry leaders, reach wider audiences on social media, and maximise their entrepreneurial success.

The researchers encourage complementing this study with qualitative research through interviews or focus groups, as this could provide further insights on specific traits, behaviours, and content that customers judge to be effective in establishing authenticity, credibility, and attractiveness in personal branding. In addition, qualitative research can help identify other variables of interest that customers might value for strengthening their connection, trust, and loyalty to the brand. Conducting a mixed-method approach could offer a more comprehensive understanding of the subject. Even more so, by evaluating the perspectives of both the fitness coaches and the customers. Future studies could also conduct a comparative analysis between personal brands on social media to encourage participants to assess the brands on the study's variables: authenticity, attractiveness, and credibility, which will offer further insights into the interplay among these variables.

References

Angel, P., Jenkins, A. and Stephens, A. (2018) 'Understanding entrepreneurial success: A phenomenographic approach', *International Small Business Journal: Researching Entrepreneurship*, 36(6), pp. 611–636. Available at: https://doi.org/10.1177/0266242618768662

Aron, A., Aron, E.N. and Norman, C. (2022) 'Self-expansion model of motivation and cognition in close relationships and beyond', in *Self and social identity*. Routledge, pp. 59–82.

Aron, A., Lewandowski, G.W., Mashek, D. and Aron, E.N. (2013) 'The self-expansion model of motivation and cognition in close relationships', in Simpson, J.A. and Campbell, L. (eds.) *The Oxford handbook of close relationships*. Oxford University Press, pp. 90–115.

Campagna, C.L., Donthu, N. and Yoo, B. (2022) 'Brand authenticity: Literature review, comprehensive definition, and an amalgamated scale', *The Journal of Marketing Theory and Practice*, 31(2), pp. 129–145. Available at: https://doi.org/10.1080/10696679.2021.2018937

Carlson, B.D. and Donavan, T. (2013) 'Human brands in sport: Athlete brand personality and identification', *Journal of Sport Management*, 27(3), pp. 193–206. Available at: https://doi.org/10.1123/jsm.27.3.193

Chen, C.P. (2013) 'Exploring personal branding on YouTube', *Journal of Internet Commerce*, 12(4), pp. 332–347. Available at: https://doi.org/10.1080/15332861.2013.859041

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. 2nd edn. Lawrence Erlbaum Associates.

Cohen, J. (1992) 'A power primer', *Psychological Bulletin*, 112(1), pp. 155–159. Available at: https://doi.org/10.1037/0033-2909.112.1.155

Connelly, B.L., Certo, S.T., Ireland, R.D. and Reutzel, C.R. (2011) 'Signaling theory: A review and assessment', *Journal of Management*, 37(1), pp. 39–67. Available at: https://doi.org/10.1177/0149206310388419

Dašić, D., Ratković, M. and Pavlović, M. (2021) 'Commercial aspects of personal branding of athletes on social networks', *Marketing*, 52(2), pp. 118–131. Available at: https://doi.org/10.5937/mkng2102118

Gehl, R. (2011) 'Ladders, samurai, and blue collars: Personal branding in Web 2.0', *First Monday*, 16(9). Available at: https://doi.org/10.5210/fm.v16i9.3579

Gorbatov, S., Khapova, S.N. and Lysova, E.I. (2019) 'Get noticed to get ahead: The impact of personal branding on career success', *Frontiers in Psychology*, 10, Article 2662. Available at: https://doi.org/10.3389/fpsyg.2019.02662

Hutson, D.J. (2013) '"Your body is your business card": Bodily capital and health authority in the fitness industry', *Social Science & Medicine*, 90, pp. 63–71. Available at: https://doi.org/10.1016/j.socscimed.2013.05.003

Kerviler, G. and Rodriguez, C.M. (2019) 'Luxury brand experiences and relationship quality for Millennials: The role of self-expansion', *Journal of Business Research*, 102, pp. 250–262. Available at: https://doi.org/10.1016/j.jbusres.2019.01.046

Kim, M. (2022) 'How can I be as attractive as a fitness YouTuber in the era of COVID-19? The impact of digital attributes on flow experience, satisfaction, and behavioral intention', *Journal of Retailing and Consumer Services*, 64, Article 102778. Available at: https://doi.org/10.1016/j.jretconser.2021.102778

Kowalczyk, C.M. and Pounders, K.R. (2016) 'Transforming celebrities through social media: The role of authenticity and emotional attachment', *Journal of Product & Brand Management*, 25(4), pp. 345–356. Available at: https://doi.org/10.1108/JPBM-09-2015-0969

Kristiansen, E. and Williams, A. (2015) 'Communicating the athlete as a brand: An examination of LPGA star Suzann Pettersen', *International Journal of Sport Communication*, 8(3), pp. 371–388. Available at: https://doi.org/10.1123/ijsc.2015-0066

Kucharska, W., Confente, I. and Brunetti, F. (2020) 'The power of personal brand authenticity and identification: Top celebrity players' contribution to loyalty toward football', *Journal of Product & Brand Management*, 29(6), pp. 815–830. Available at: https://doi.org/10.1108/JPBM-02-2019-2241

Kulkarni, P. (2024) '5 biggest challenges of online coaching', *LinkedIn*, 22 November. Available at: https://www.linkedin.com/pulse/5-biggest-challenges-online-coaching-pramod-kulkarni-gdhjc

Lee, J.A. and Eastin, M.S. (2021) 'Perceived authenticity of social media influencers: Scale development and validation', *Journal of Research in Interactive Marketing*, 15(4), pp. 822–841. Available at: https://doi.org/10.1108/JRIM-12-2020-0253

Li, W., Ding, H., Xu, G. and Yang, J. (2023) 'The impact of fitness influencers on a social media platform on exercise intention during the COVID-19 pandemic: The role of parasocial relationships', *International Journal of Environmental Research and Public Health*, 20(2), Article 1113. Available at: https://doi.org/10.3390/ijerph20021113

Lunardo, R., Gergaud, O. and Livat, F. (2015) 'Celebrities as human brands: An investigation of the effects of personality and time on celebrities' appeal', *Journal of Marketing Management*, 31(5–6), pp. 685–712. Available at: https://doi.org/10.1080/0267257X.2015.1008548

Mauri, A.G., Minazzi, R., Nieto-García, M. and Viglia, G. (2018) 'Humanize your business: The role of personal reputation in the sharing economy', *International Journal of Hospitality Management*, 73, pp. 36–43. Available at: https://doi.org/10.1016/j.ijhm.2018.01.017

Mintel (2024) *UK health and fitness clubs market report 2024*. Available at: https://store.mintel.com/report/uk-health-and-fitness-clubs-market-report

Mousavi-Jad, S.M., Shafei, R. and Emami, E. (2021) 'An analysis of personal branding tactics strategies in athletes' professional success', *Journal of International Marketing Modeling*, 2(1), pp. 31–40.

North, J., Piggott, D., Rankin-Wright, A. and Ashford, M. (2020) 'An empirical examination of U.K. coaches' issues and problems, and their support and advice networks', *International Sport Coaching Journal*, 7(3), pp. 283–294. Available at: https://doi.org/10.1123/iscj.2019-0049

Office for National Statistics (2024) *Self-employment in the UK*. Available at: https://www.ons.gov.uk

Ohanian, R. (1990) 'Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness', *Journal of Advertising*, 19(3), pp. 39–52. Available at: https://doi.org/10.1080/00913367.1990.10673191

Phung, L. and Qin, L. (2018) *Perception of social media influencers: A study on evaluation of social media influencer types for different beauty categories*. Master's thesis, Jönköping University. Available at: http://hj.diva-portal.org/smash/record.jsf?pid=diva2%3A1213878

Pontinha, V.M. and Coelho do Vale, R. (2020) 'Brand love measurement scale development: An inter-cultural analysis', *Journal of Product & Brand Management*, 29(4), pp. 471–489. Available at: https://doi.org/10.1108/JPBM-10-2018-2094

Reimann, M., Castaño, R., Zaichkowsky, J. and Bechara, A. (2012) 'How we relate to brands: Psychological and neurophysiological insights into consumer–brand relationships', *Journal of Consumer Psychology*, 22(1), pp. 128–142. Available at: https://doi.org/10.1016/j.jcps.2011.11.003

Resnick, S.M., Cheng, R., Simpson, M. and Lourenço, F. (2016) 'Marketing in SMEs: A "4Ps" self-branding model', *International Journal of Entrepreneurial Behavior & Research*, 22(1), pp. 155–174. Available at: https://doi.org/10.1108/IJEBR-07-2014-0139

Sakib, M.N., Zolfagharian, M. and Yazdanparast, A. (2020) 'Does parasocial interaction with weight loss vloggers affect compliance? The role of vlogger characteristics, consumer readiness, and health consciousness', *Journal of Retailing and Consumer Services*, 52, Article 101867. Available at: https://doi.org/10.1016/j.jretconser.2019.01.002

Scheidt, S., Gelhard, C. and Henseler, J. (2020) 'Old practice, but young research field: A systematic bibliographic review of personal branding', *Frontiers in Psychology*, 11, Article 1809. Available at: https://doi.org/10.3389/fpsyg.2020.01809

Scheidt, S. and Henseler, J. (2018) 'Personal branding: A review on a contemporary phenomenon', *7th DERMARKENTAG*, Germany, 15–16 November. Available at: https://www.researchgate.net/publication/346914136

Shepherd, I.D.H. (2005) 'From cattle and coke to Charlie: Meeting the challenge of self-marketing and personal branding', *Journal of Marketing Management*, 21(5–6), pp. 589–606. Available at: https://doi.org/10.1362/0267257054933469

Spence, M. (1973) 'Job market signaling', *The Quarterly Journal of Economics*, 87(3), pp. 355–374. Available at: https://doi.org/10.2307/1882010

Staškevičiūtė-Butienė, I., Bradauskienė, K. and Crespo-Hervas, J. (2014) 'Athletes' personal brand as a success factor for start-up', *Transformations in Business and Economics*, 13(2A), pp. 526–540.

Statista (2023) *Fitness industry in the United Kingdom (UK)*. Available at: https://www.statista.com/study/39583/fitness-industry-in-the-united-kingdom-uk-statista-dossier/

Thai, N.T. and Wang, J. (2020) 'The effect of brand authenticity and brand attachment on consumer product-sharing', *Journal of Strategic Marketing*, 28(7), pp. 653–666. Available at: https://doi.org/10.1080/0965254X.2019.1567571

Thompson-Whiteside, H., Turnbull, S. and Howe-Walsh, L. (2018) 'Developing an authentic personal brand using impression management behaviours', *Qualitative Market Research: An International Journal*, 21(2), pp. 166–181. Available at: https://doi.org/10.1108/QMR-01-2017-0007

Vitelar, A. (2019) 'Like me: Generation Z and the use of social media for personal branding', *Management Dynamics in the Knowledge Economy*, 7(2), pp. 257–268. Available at: https://doi.org/10.25019/MDKE/7.2.07

Wei, Z., Zhang, M. and Qiao, T. (2022) 'Effect of personal branding stereotypes on user engagement on short-video platforms', *Journal of Retailing and Consumer Services*, 69, Article 103121. Available at: https://doi.org/10.1016/j.jretconser.2022.103121

Zhou, F., Mou, J., Su, Q. and Jim Wu, Y.C. (2020) 'How does consumers' perception of sports stars' personal brand promote consumers' brand love? A mediation model of global brand equity', *Journal of Retailing and Consumer Services*, 54, Article 102012. Available at: https://doi.org/10.1016/j.jretconser.2019.102012

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 55-74
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1738

Risk in Outsourced IT Operations: A Systematic Literature Review of Technological Uncertainty, Knowledge Management and Opportunistic Behaviour

[1]Peter Hewes, [2]Marcelo Martins Sa, [3]Adrian Small, [4]Dongjun Li
Northumbria University

Abstract
Technological adoption and digital transformation are increasingly enabling organisations to gain a competitive advantage, underpin business processes and create efficiencies. Utilisation of technologies brings several risks such as the risk of cyber-attacks and infrastructure dependency. To mitigate these, and to reap other benefits, organisations are increasingly turning to third-party IT suppliers to innovate and manage their IT estates. The field of IT operations and supply chain management has gained extensive research over the decades; however, none seem to bring these aspects together with technology risk mitigation in a systematic way. This paper systematically reviews 63 articles to ascertain historical research trends to indicate future research interest and consolidate research themes to discuss the gaps in the extant research. Alongside the fact that academic interest is projected to continue, closely coupled with world events, important findings show that technological uncertainty, information asymmetry and opportunistic behaviour are closely coupled in outsourced IT operations, and that knowledge management acts as a key mitigation mechanism which is illustrated by a new conceptual model. The review also reveals that existing research focuses heavily on ex-ante IT operations outsourcing decisions, with limited attention given to the ex-post operational phase, where most risks in IT operations materialise. Several gaps are identified for the field including how knowledge management can be utilised to mitigate technology risk within the IT operations function linking out to technology adoption. More generally, research into the public sector is found to be underreported giving researchers another lens to investigate current research themes with or adopt those previously listed. Overall, the review provides an integrated understanding of technological uncertainty in outsourced IT operations and highlights key opportunities for further research into ex-post phase, specifically long-term knowledge management in sector-specific outsourcing.

Keywords: IT Operations; IT Outsourcing; Knowledge Management; Technology Risk; Technology Adoption; Systematic Literature Review (SLR)
Wordcount: 286

1.0 Introduction

It is widely accepted that technology is advancing at pace, enabling businesses to undergo digital transformation and technological adoption to support, underpin and enhance business processes. With that, the need and reliance on third parties to supply technological services is also increasing (Pacheco and Paul, 2023). With cost being the main driver for utilising third-party suppliers to innovate, develop and run IT services, other drivers include access to expert resources and IT competence benefits (Qu and Pinsonneault, 2011). With the above benefits to IT outsourcing come limitations, these include loss of control, opportunism and in some cases, higher transaction costs (Hansen et al., 2019, Khan et al., 2019). This can have consequences for the roles and functions undertaken by the internal IT department, such as knowledge management, and therefore decision-making capacity.

Since Kodak pioneered outsourcing the IT business process to a third-party supplier in 1989, academic interest in the subject, in line with general supply chain management has grown. Adding to this, several world events over the past few decades such as the Y2K bug, the global financial crisis of 2008 and the COVID-19 pandemic have changed the landscape for IT, forcing organisations to adapt, digitise and change sourcing strategies, maintaining academic interest in the area. Despite the growing literature, this body of research tends to focus on only one dimension at a time, for example, outsourcing governance, Transaction Cost Economics, or technological risk as evidenced in prior SLR's examining only isolated aspects (Lacity and Willcocks, 2014, Lin and Vaia, 2015, Nduwimfura and Zheng, 2015). Only when integrating these aspects will researchers be able to explain how organisations can mitigate risk once IT operations have been outsourced. Due to further limited clarity on what constitutes as "IT Operations" for the purposes of this review, this refers specifically to the ongoing run, maintenance and evolution of infrastructure, applications and services which is distinct from software development outsourcing or project management. This falls in line with more common IT outsourcing literature, which focuses more on ex-ante sourcing decisions rather than the ex-post operational phase where technological uncertainty, information asymmetry and opportunistic behaviour can most strongly affect organisational outcomes (Hanafizadeh and Zareravasan, 2020).

Only by contrasting this viewpoint, does the need for this SLR become necessary. By reviewing all literature covering these aspects since the first papers began to emerge in 1992, this study aims to consolidate the position so far, enabling the development of both academic inquiry and real-world decision making. Following the six-step process proposed by Sauer and Seuring (2023) and illustrated in Figure 1, this SLR provides a modern baseline for future research to build upon and identifies both generalised and specific research directions that scholars may wish to pursue. One key area of future interest is understanding the gap in public sector reporting within the field, not only technically how public sector organisation IT operations operate, but what barriers there are to research and publication in this field specifically.

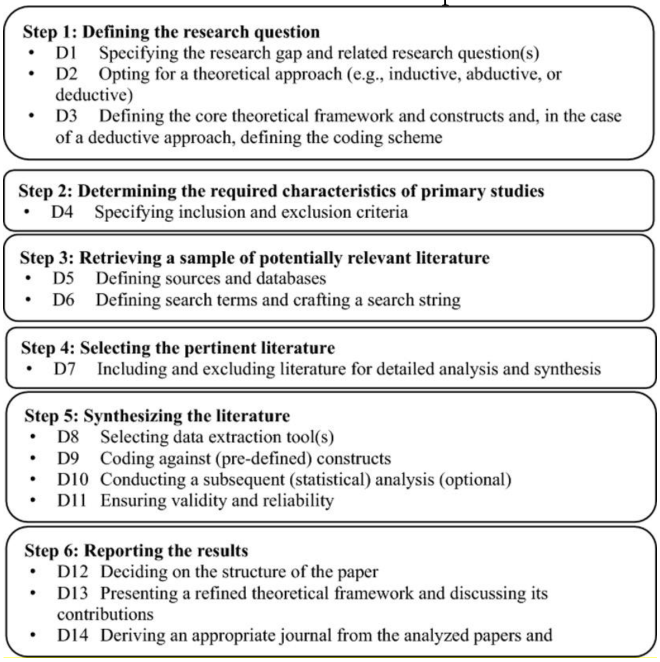

Figure 1 Six step process to undertaking a SLR Sauer and Seuring (2023)

This SLR advances knowledge in IT Operations and Supply Chain Management in the following ways:

- Identifies research gaps and future direction in the field of IT Operations Management specifically concerning how knowledge management can affect various uncertainties.
- Validates a modern approach to undertaking an SLR as outlined by Sauer and Seuring (2023).
- Analyses research trends and themes within the field to indicate future research interest.

The paper is set out as follows: section 2 outlines the methodology used defining the decision points for the SLR and why those decisions were taken. Section 3 contains the results and section 4 the discussion, outlining the findings of the SLR and directly addresses the research questions set out in the paper. Section 5 proposes the future research directions, and the paper concludes with section 6 where final summaries are made.

2.0 Methodology

This systematic literature review (SLR) aims to investigate the operational management of information technology (IT) supply chains by using knowledge management and the considerations this brings to risk management.

To achieve this, the framework to undertaking a SLR as outlined by Sauer and Seuring (2023) has been followed meaning the process is divided into a six-step process split further by fourteen decision points outlined in the following sections.

2.1. *Defining the Research Question*

This step sees the outline of the research questions at decision point one. The research questions are formulated through the proposed review topic outlined above, to holistically review the extant literature and establish where the literature could be updated in the future.

- RQ1: What are the trends in the extant research?
- RQ1.1: What are the publishing trends surrounding the management of IT in an outsourced model?
- RQ1.2: What types of methodologies are used to evaluate research surrounding the management of IT in an outsourced model?
- RQ1.3: In which countries are research surrounding the management of IT in an outsourced model undertaken?
- RQ1.4: What sectors have attracted the most research attention?
- RQ1.5: What are the key theoretical frameworks used to assess the research area?
- RQ2: What are the main themes surrounding the management of IT in an outsourced model?
- RQ3: What are the gaps in the extant research for future research direction?

The second decision point is the identification and adoption of a theoretical approach. For this SLR, an inductive theoretical approach was adopted. This is because it allows theory identification through the analysis of patterns and themes of the mainly qualitative data gathered on a more general topic of focus (Sabherwal and King, 1991) as opposed to a deductive approach, which begins with a hypothesis to either prove or falsify often through quantitative means such as experiments (Varpio et al., 2020). This is not to say that a deductive approach to proving the theoretical output of this inductive SLR cannot be used (Hyde, 2000).

The final decision point for step one is to identify a theoretical framework. The approach outlined by Sauer and Seuring (2023) does not explicitly require the adoption of a theoretical framework at step one, and in fact, they do not in their own SLR used to define the process. So, given the exploratory nature of this review, a theoretical lens is not imposed. This allows for a broader and more inclusive identification of relevant literature, avoiding a premature narrowing of scope.

2.2. *Determining the characteristics of primary studies*

Using guidelines outlined by Pilbeam et al. (2012) and Durach et al. (2017), the inclusion and exclusion criteria (outlined in Table 1) makes up the outcome of decision four.

Criteria	Include/Exclude	Rationale
Publication in peer-reviewed journals	Include	It is expected that peer-reviewed articles are of higher quality than other documents.
Contain a main theme of any of the research themes with respect to outsourcing; IT Operations, Knowledge Management, Technology Risk, Opportunistic Behaviour or Technology Adoption	Include	The focus of the SLR is around outsourcing IT and the supply chain that surrounds that. These are the key themes linked to RBV.
Paper may be published in any year	Include	The SLR is inclusive of all publications to provide a holistic view.
Studies utilising qualitative and quantitative methodologies presenting empirical, theoretical or literature reviews	Include	All approaches to research have contributed to extant literature.
Papers relating to topics which do not focus on outsourcing	Exclude	The main theme of the SLR is outsourcing.
Papers in any language other than English	Exclude	The author is only able to read English.
Papers to which there is no access to the full text	Exclude	The full text must be available to be reviewed.

Table 1 SLR inclusion/exclusion criteria

This decision point concludes the second step in the SLR process and allows a criterion for the sample to be assessed against to provide a baselined sample.

2.3. *Retrieving a sample of potentially relevant literature*

Upon defining the above criteria, it is then possible to retrieve a sample of literature from online databases. To do so, first the target databases must be identified, this is decision four. Although as outlined by Pearce (2018), there are many free databases sources for use in collecting samples, for this SLR, it was decided to collect data from both Scopus and Web of Science (WoS) databases as these are available via institutional access and are regarded the most globally influential citation indexes (Asubiaro et al., 2024). Another reason for this is because some sources are not available in both databases, so in doing this, it is possible to obtain a more complete dataset.

The next step in the SLR process is to retrieve the dataset by generating search strings, this makes up decision six. To do this, several keywords were identified relating to the general topic outlined in section 1. Some keyword examples were "opportunistic behaviour", "information technology" and "ITO" to ensure abbreviations were used as well as the keywords themselves.

All search strings were run against both databases on the same day (30th August 2023) to ensure consistency and reliability providing a total of 142 results. Because the keyword list was intentionally broad, the search strategy carries an inherent risk of conceptual drift (e.g., pulling in technology-risk papers unrelated to outsourcing). To mitigate this, thematic relevance was assessed at both title/abstract and full-text stages, and the rationale for exclusion of borderline items was recorded as described in section 2.4. This reflection on search breadth and potential bias increases methodological transparency.

2.4. *Selecting the pertinent literature*

Step four of the SLR process as outlined by Sauer and Seuring (2023), is to select the sample data used in the review. The 142 output documents generated from the search strings being ran against the two databases underwent a selection process outlined by Moher et al. (2009) and illustrated as a Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) diagram in Figure 2. Step four of this process neatly overlays the screening, eligibility and inclusion phases outlined by Moher et al. (2009). To automate the removal of duplicate documents output by numerous search strings against the two databases, the Bibliometrix package for RStudio was used (Aria and Cuccurullo, 2017). This utilised code to amalgamate all documents, remove duplicates and generate a matrix to be used in the subsequent steps of the screening phase. This resulted in the removal of 27 documents. The output matrix included various bits of information including a language and DOI allowing for a fast exclusion of non-English documents and where the full text was not available via the home institution as per exclusion criteria set out in Table 1. The resulting list of 97 full texts were then tested against the remaining inclusion/exclusion criteria providing a final sample of 63 documents for review.

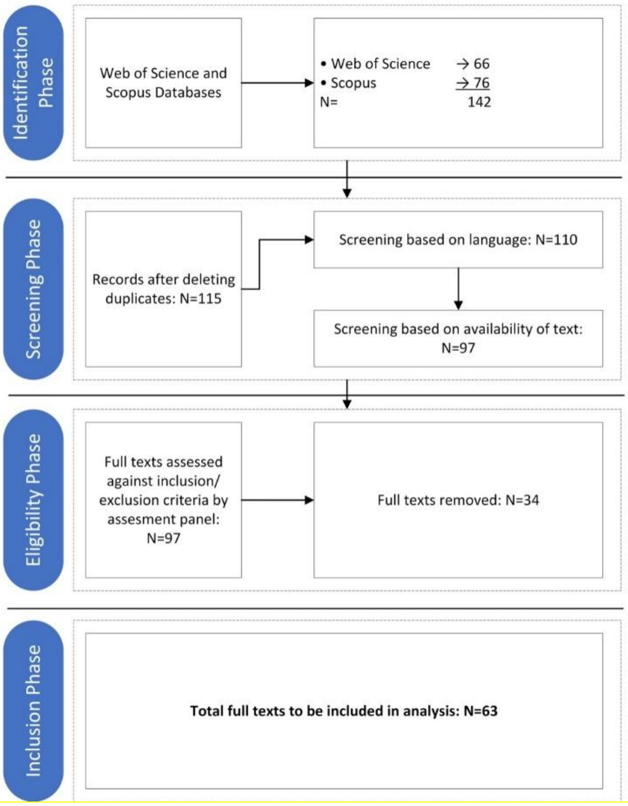

Figure 2 Visualisation of Step Four (Selecting the Pertinent Literature)

To enhance reliability, a secondary reviewer independently screened the papers and reviewed all borderline cases with the author at review meetings. Disagreements were resolved through discussion, and where uncertainties persisted, the more inclusive option was chosen. Where relevance was unclear, cases were retained for full-text review. Borderline cases such as papers addressing technology risk without an explicit outsourcing component were discussed jointly with the secondary reviewer to ensure consistent interpretation of thematic relevance. This helped minimise the risk of unintentionally excluding papers where outsourcing was embedded but thematic relevance not highlighted. Although a formal inter-rater statistic (Cohen's kappa) was not utilised, this collaborative approach helped ensure consistent application of inclusion criteria.

Additionally, during this stage a short reflective log was maintained documenting reasons for exclusion of borderline papers and used for prompt during review meetings. This provided an audit trail supporting the transparency and reproducibility expected of SLRs at Q1 standard. A methodological note must also be made to acknowledge the limits of access-based limitations which could introduce publication bias.

2.5. Synthesising the literature

The first decision of step 5, decision eight, is to decide on the data extraction tool to be used to code and synthesise the data. QDAS programmes are heavily recommended for this by Sauer and Seuring (2023) due to their impressive functionality to store documents, auto and manual code, while offering tabulated comparison and comprehensive qualitative analysis and querying. More specifically, O'Neill et al. (2018) outline on how NVivo can be used to code text, link them and generate visualisations and relationships to support a SLR. This approach offers more functionally than other spreadsheet-based options given by Sauer and Seuring (2023), and although Microsoft Excel is the most commonly used tool for synthesising SLR literature, it is not specifically designed for this functionality (Al-Zubidy and Carver, 2019). For these reasons, NVivo was the tool chosen to undertake the synthesis of the sample due to its ability to systematically store coded excerpts and run cross-case queries. As a coding protocol was established prior to analysis, this reduced researcher subjectivity and aligns with recommended SLR practice.

To fairly code the sample, it is important to pre-determine the coding criteria. This ensures that as the SLR progresses coding structure is well documented, reliable and understood (Sauer and Seuring, 2023). On this premise, the main themes were set out as parent codes (Outsourced IT Operations, Knowledge Management, Technology Risk, Opportunistic Behaviour or Technology Adoption) as outlined in section 1. These parent codes then contained multiple sub-codes, for example, 'Barriers to Technology Adoption' as a sub-code of 'Technology Adoption'. Other codes were used to help analyse literature trends such as research methodology, journal, year to enable analysis on where the literature has been and to identify gaps or areas for future work.

Table 2 outlines the sources forming the final dataset for the SLR. The table shows where there was relevance to one or more of the main themes and were coded during the synthesis part of the SLR process.

Categories	Authors
Technology Risk	(Lowis and Accorsi, 2011), (Worrell et al., 2013), (Tang et al., 2018), (Samtani et al., 2020), (Jacobs et al., 2020), (Mantha and García de Soto, 2021), (Leverett et al., 2022), (Baho and Abawajy, 2023), (Lin et al., 2023)
Outsourced IT Operations, Opportunistic Behaviour	(Lin and Hekkala, 2016), (Qin et al., 2018), (Awe et al., 2018), (Hansen et al., 2019), (Nepomuceno et al., 2022)
Knowledge Management, Opportunistic Behaviour	(Aron et al., 2005), (Fauchart, 2006), (Kloyer and Scholderer, 2012), (Yang et al., 2020)
Outsourced IT Operations, Knowledge Management, Technology Risk, Opportunistic Behaviour	(Frigant, 2011), (Qu and Pinsonneault, 2011), (Mathew and Chen, 2013), (Khan et al., 2019)
Outsourced IT Operations, Technology Risk, Opportunistic Behaviour	(Kauffman and Tsai, 2009), (Handley and Benton, 2012), (Hansen et al., 2017), (Haq et al., 2019)
Outsourced IT Operations, Knowledge Management, Opportunistic Behaviour	(Rustagi et al., 2008), (Cruz and Liu, 2011), (Cheng and Chen, 2016), (Hoseini et al., 2020)
Knowledge Management, Technology Risk, Opportunistic Behaviour	(Szczepański and Świątowiec-Szczepańska, 2012), (Adams and Graham, 2017), (Yang et al., 2021)
Knowledge Management, Technology Risk	(Lean and Tucker, 2005), (Palomino et al., 2013), (Syed, 2020)
Knowledge Management	(Al-Karaghouli et al., 2005), (Meenan et al., 2010), (Mishra et al., 2024)
Outsourced IT Operations, Knowledge Management, Technology Risk, Opportunistic Behaviour, Technology Adoption	(Schneider and Sunyaev, 2016), (Öbrand et al., 2019)
Outsourced IT Operations, Technology Risk, Opportunistic Behaviour, Technology Adoption	(Yigitbasioglu et al., 2013), (Chang et al., 2017)
Knowledge Management, Technology Risk, Opportunistic Behaviour, Technology Adoption	(Ravichandran, 2005), (Hong et al., 2010)
Outsourced IT Operations, Knowledge Management, Opportunistic Behaviour, Technology Adoption	(Bui et al., 2019), (Acharya et al., 2022)
Outsourced IT Operations, Technology Risk, Technology Adoption	(Marinč, 2013), (Eachempati, 2017)
Knowledge Management, Technology Adoption	(Eze et al., 2019), (Pacheco and Paul, 2023)
Outsourced IT Operations, Technology Risk	(Aubert et al., 2012), (Biswas and Mukhopadhyay, 2018)
Technology Risk, Opportunistic Behaviour	(Clemons and Kleindorfer, 1992), (Hoffmann et al., 2013)
Outsourced IT Operations, Knowledge Management, Technology Risk, Technology Adoption	(Ali et al., 2022)
Technology Risk, Opportunistic Behaviour, Technology Adoption	(Silic and Back, 2016)
Knowledge Management, Opportunistic Behaviour, Technology Adoption	(Pistoni et al., 2022)
Outsourced IT Operations, Knowledge Management	(Refaiy and Labib, 2009)
Outsourced IT Operations	(Wilkin et al., 2016)
Opportunistic Behaviour	(DuHadway et al., 2022)
None	(Valiente et al., 2012)

Table 2 Sources included in final dataset.

2.6. *Reporting the results*

The final step in the framework is to report the results; to complete this, statistical analysis was conducted on the output coded matrix extracted from NVivo. This allowed the generation of various charts to report on thematic trends. To complete the discussion, NVivo was used to analyse codes reported on throughout the process.

The results and subsequent discussion follow in the next section.

3.0. Results

This section provides an overview of the results of the SLR using statistical analysis and illustration where necessary to directly address the research questions presented in section 2.1.

3.1. RQ1: What are the trends in the extant research?

RQ1.1: What are the publishing trends surrounding the management of IT in an outsourced model?

There are several pieces of analysis that can depict the publishing trends around the research area as outlined in work undertaken through various reviews (Baho and Abawajy, 2023). Figure 3 illustrates the trend in the publications distributed by year. The result shows a general upward trend in the number of publications per year.

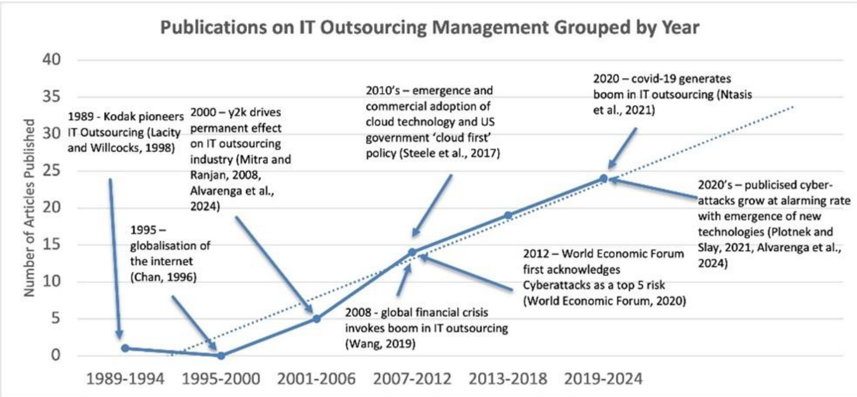

Figure 3 Chart illustrating publications related to IT outsourcing management distributed by year

The growing upward trend in publication indicates that increasing attention is being paid to the general topic of management of IT operations where outsourcing is involved. This could be due to the ever-evolving nature of technology and ramifications of events requiring the use of technology as a solution. Figure 3 overlays a few significant events which have impacted the general field and these in turn could have influenced the interest. A forecast line is also plotted indicating that the interest in this field is predicted to further increase. Figure 4 illustrates the top journals within the dataset. This gives an indication of the top 10 journals in which the dataset sources are published in. In total, the entire dataset contained articles published in 52 journals, and the remaining journals contained only 1 published article per journal.

Figure 4 Chart illustrating the most relevant sources.

Table 3 outlines the CABS academic journal ranking (CABS, 2021) and SJR (SJR, 2024) H-index for the most relevant journals. This shows that the entirety is internationally recognised (2 star) with the majority being either 3 star (internationally excellent) or 4 stars (world leading).

Journal	H-Index	Ranking	No. of Documents
Journal of Strategic Information Systems	107	4	3
Journal of Management Information Systems	170	4	2
IEEE Transactions on Engineering Management	112	3	3
Industrial Marketing Management	177	3	2
Journal of Purchasing and Supply Management	100	3	2
International Journal of Production Research	186	3	2
Decision Support Systems	180	3	1
Journal of Enterprise Information Management	82	2	2
Journal of Global Information Management (JGIM)	46	2	2
International Journal of Accounting Information Systems	65	2	2

Table 3 H-Index and CABS ranking for most relevant sources.

This indicates that there is a respected interest in the research field and positively reenforces it as a research direction.

RQ1.2: What types of methodologies are used to evaluate research surrounding the management of IT in an outsourced model?
As with any research topic forming part of a literature review, there is a plethora of research methodologies used across the dataset. In this instance, Figure 5 illustrates the spread of methodologies used in the dataset.

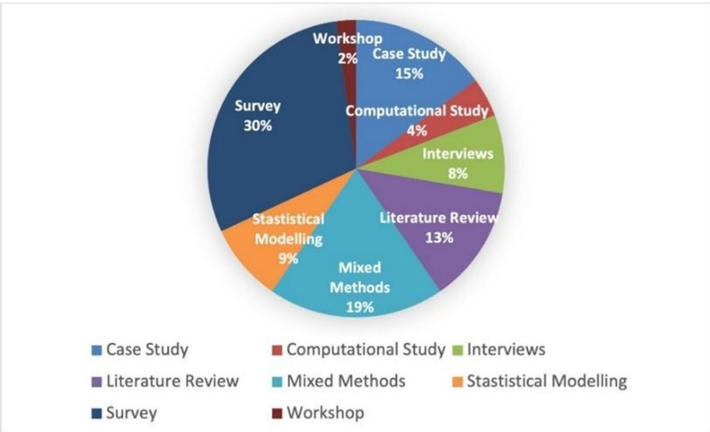

Figure 5 Chart illustrating distribution of research methodologies

As seen above, most of the methodologies used across the research are qualitative with surveys being widely used. Mixed methods are frequently used to allow for the inclusion of a quantitative element to confirm hypotheses output by an exploratory (qualitative) element such as seen in the study undertaken by Refaiy and Labib (2009) whereby postal survey data alongside interviews was used to prove that the sharing of tacit knowledge has a positive impact on maintenance performance while also taking into account organisational aspects.

RQ1.3: In which countries are research surrounding the management of IT in an outsourced model undertaken?
Coding for the country in which the primary research was undertaken, rather than the country that the authors originate from allow illustration of where the global focus resides concerning the research area of management of IT operations utilising outsourcing.

Figure 6, supported by Table 4 illustrate the distribution by research location.

Figure 6 Country heatmap illustrating research location.

Of the 63 studies, a large number (26) did not specify the country or location in which the research was carried out, a further 4 studies were self-identified as being world-wide studies while 4 concerned covering Europe in its entirety. As such, these 34 studies could not be incorporated into Figure 6 but are contained in Table 4.

Country	Number of Publications
Not Specified	26
United States	10
Europe	4
United Kingdom	4
Worldwide	4
Australia	3
Canada	2
China	1
Germany	1
India	1
Italy	1
South Korea	1
Nepal	1
Netherlands	1
Pakistan	1
Poland	1
Portugal	1
Taiwan	1
United Arab Emirates	1

Table 4 Number of publications by the location of the research

RQ1.4: What sectors have attracted the most research attention?
Like the coding structure for country, it was also possible to code for the primary research focus in terms of sector. In this instance, four codes were used; agnostic is where both private and public sector

organisations were studied but not differentiated. Private and public sector were tagged accordingly, if the primary research focused entirely on that sector, and not specified was used where the authors did not specify. This split is illustrated in Figure 7.

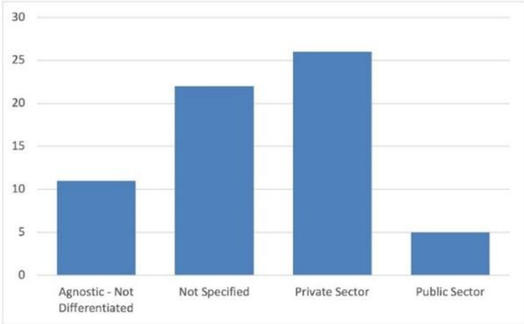

Figure 7 Chart illustrating the research split by sector

As seen above, the data indicates that most of the research either focuses primarily on private sector organisations or does not specify from which sector the research data originates. It can also be noted that of the five articles with a public sector focus, only the work undertaken by Ali et al. (2022) actually collects and uses primary data from within a public sector organisation. These numbers suggest there is a potential barrier to research in the public sector concerning outsourced IT operations management.

RQ1.5: What are the key theoretical frameworks used to assess the research area?

Like the above, it is possible to use the data to understand the trend in theoretical focus for the subject area of outsourced IT operations management. Figure 8 illustrates the split of theoretical frameworks used as a lens to the research showing the top five frameworks used. Any frameworks in the 'other' category are equally split twelve ways and each is only used as a lens once in the entire dataset. Some of these are Nash Bargaining Theory (Clemons and Kleindorfer, 1992), Reliability Theory (Qin et al., 2018) and Industry Clockspeed Theory (Kauffman and Tsai, 2009).

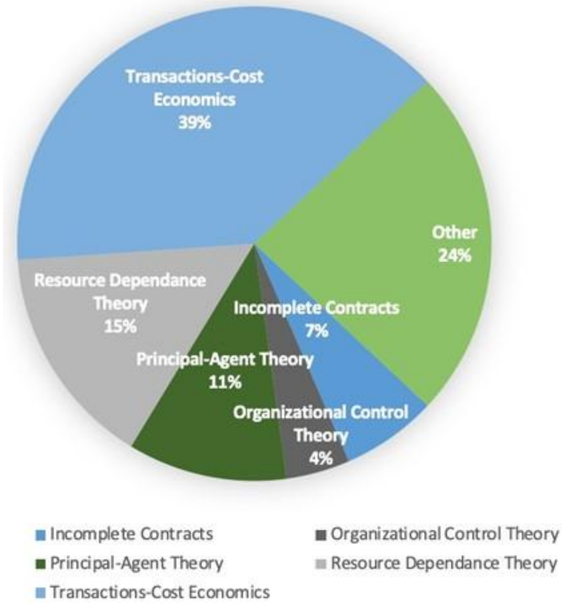

Figure 8 Chart illustrating the split of theoretical frameworks used in the dataset.

The theoretical framework which is used most widely across the dataset is Transaction Cost Economics (TCE) with the reasoning being relevance to the subject area. TCE takes costs for multiple aspects of the product or service such as negotiating costs, uncertainty and run costs to help the organisation decide the sourcing or governance structure (Aubert et al., 2012). In the dataset, a large proportion of the sources focus on using TCE as a framework to enable strategic decisions and sourcing strategies (Hansen et al., 2019, Haq et al., 2019, Khan et al., 2019) while fewer focus on the inter-organisational relationships and risks during the implementation/run phase of the contract (Chang et al., 2017, DuHadway et al., 2022). This could be an opportunity for further research in the field due to the modern adoption of generic contracts (Bui et al., 2019).

4.0. Discussion

Building on the findings presented in the results section, this section critically examines the key themes emerging from the literature surrounding IT operations in outsourced models. By synthesising insights across the dataset, the section explores how organisations navigate challenges such as opportunistic behaviour, technology risk, and knowledge management, and how these intersect with technology adoption.

RQ2: What are the main Research themes surrounding the management of IT operations in an outsourced model?

Upon analysing the articles contained in the dataset it is possible to answer RQ2 by outlining the main themes surrounding the main topic outlined in section to summarise extant research and identify research gaps to feed in to RQ3. Attention is paid to how these themes interact rather than treating them in isolation, reflecting the interconnected nature of outsourced IT operations.

IT Operations

IT operations encompass all the activities and responsibilities involved in managing and maintaining an organisation's IT infrastructure and services. As with other business processes, multiple sourcing strategies can be adopted to undertake IT operations on behalf of a company including a fully IT Outsourcing (ITO) model which is defined by Schneider and Sunyaev (2016, p.3) as "the significant contribution by external vendors in the physical and/or human resources associated with the entire or specific components of the IT infrastructure in the user organization"

Keeping in line with general outsourcing literature, many authors identify the decision to outsource IT operations is largely based on costs, particularly labour and production (Aubert et al., 2012, Mathew and Chen, 2013, Yigitbasioglu et al., 2013), however, some authors have challenged the above, suggesting that insourcing IT operations could better meet user requirements (Ravichandran, 2005), retain knowledge about assets or the IT estate (Qu and Pinsonneault, 2011, Schneider and Sunyaev, 2016) or simply avoid inflated transaction costs and opportunistic behaviour (Handley and Benton, 2012, Öbrand et al., 2019).

Across the dataset, it becomes clear that the decision to outsource is tightly coupled with downstream management challenges such as cost and organisational capability, this is not a new concept in supply chain literature. However, organisations often underestimate the capability required to effectively manage outsourced operations, creating conditions in which technological uncertainty and opportunistic behaviour become more likely. This shows that IT operations form the backbone upon which all other themes (risk, knowledge and opportunism) interact.

Opportunistic Behaviour

Throughout the dataset, several types of opportunistic behaviour are identified such as shirking (Aron et al., 2005, Handley and Benton, 2012, Mathew and Chen, 2013), poaching (Aron et al., 2005) and the withholding of information (Kauffman and Tsai, 2009, Mathew and Chen, 2013, Haq et al., 2019), all of which can occur ex-ante (pre-contract) or ex-post (post-contract) (Aron et al., 2005, Kloyer and Scholderer, 2012, Qin et al., 2018, Khan et al., 2019). There is a limitation of research focus throughout the dataset, where IT operations is concerned whereby most articles focus on the decision to outsource (ex-ante) either IT operations as a business process, or an IT software development project with little focus given to ex-post management of suppliers and opportunistic behaviour for IT operations.

Causes of opportunistic behaviour are commonly outlined during the introduction or literature review section of the articles included in the dataset often acknowledging issues such as information asymmetry (Fauchart, 2006, Nepomuceno et al., 2022, Pistoni et al., 2022), vendor lock-in (Handley and Benton, 2012, Chang et al., 2017) and the uncertainty of the IT industry (technological risk) (Ravichandran, 2005, Khan et al., 2019, Yang et al., 2020).

The above types and causes of opportunistic behaviour are addressed by mitigation methods such as monitoring and control changes (Aron et al., 2005, Fauchart, 2006, Rustagi et al., 2008), better contracts or financial incentives (Kauffman and Tsai, 2009, Frigant, 2011), and improved knowledge capability of the client firm (Mathew and Chen, 2013, Yigitbasioglu et al., 2013, Adams and Graham, 2017). Although these mitigation methods are confirmed as effective, they are generalisations not specific to the unique IT industry where technological innovation and uncertainty is extremely high (Lean and Tucker, 2005). With that, modern ways of working are not considered in recent literature and specific mitigations such as organisational design, and knowledge strategy are not prescribed in a modern way.

Building on the above, the key insight emerging from this review is that opportunistic behaviour is amplified when organisations lack the knowledge capabilities necessary to interpret supplier actions. Therefore, knowledge asymmetry is not just a product of opportunism, it is also an enabler, creating a cycle. This highlights the importance of examining opportunistic behaviour not in isolation, but as dynamically linked to both technological uncertainty and knowledge management competence.

Technology Risk

In today's digital world, technology risks not only impact the IT infrastructure of an organisation, but can have negative effects on the business processes they enable (Worrell et al., 2013). There are many different flavours of technology risk such as cyber-security risks (Tang et al., 2018), the speed at which the industry is evolving (Lean and Tucker, 2005, Ravichandran, 2005, Frigant, 2011, Qu and Pinsonneault, 2011) and finally industry capacity and capability limitations (Hong et al., 2010). Lowis and Accorsi (2011) find that most systems are exploited through widely known vulnerabilities despite patches being available. This is supported by Jacobs et al. (2020) who found that firms have more vulnerabilities than resources to fix during their study to find improvements in exploit prediction which could put them at further risk of higher transaction costs or exploitation by an opportunistic IT supplier (Hong et al., 2010). Chang et al. (2017) argue, however, that it is impossible to fully specify IT contracts due to technological risk which indicates a risk management gap between the polars, one which is not reported on in the dataset.

Mitigations to technology risk are largely reported in the dataset to be the management of vulnerability databases such as The Common Vulnerability Exposure (CVE) and the National Vulnerability Database (NVD) (Syed, 2020, Leverett et al., 2022, Lin et al., 2023). To do this, however, firms must increase investment in technology and alter procurement strategies when outsourcing their IT operations (Kauffman and Tsai, 2009, Qu and Pinsonneault, 2011). Most of the dataset focuses on the private sector response to this, and does not stipulate how this could be undertaken best in the public sector where budgets are low and politically driven, and where goals are to reduce IT operations and infrastructure costs (Lean and Tucker, 2005, Ali et al., 2022). Building on this, Mantha and García de Soto (2021) find that cyber risk is intensified on construction projects where public safety is concerned further magnifying the gap in research.

Importantly, this review reveals that technological uncertainty intensifies relational risks. When organisations do not fully understand emerging technologies, or the technicalities/interdependencies of their IT estate, they become more reliant on suppliers. This directly drives exposure to the knowledge asymmetry cycle as described above, demonstrating a bidirectional relationship where technology risk feeds information asymmetry, which in turn exacerbates relational risk of opportunistic behaviour.

Technology Adoption

Technology adoption of an organisation is somewhat underreported throughout the dataset, the focus is on transformative adoption such as the adoption of cloud computing (Marinč, 2013, Schneider and Sunyaev, 2016). Other authors are indicating that the technology procurement strategies of organisations and suppliers is changing with a shift to transactional licencing models of Software as a Service (SaaS) and Open Source Software (OSS) from traditional bespoke build strategies with suppliers becoming more pre-emptive in their offerings of services (Yigitbasioglu et al., 2013, Silic and Back, 2016).

Ali et al. (2022) find that there is a direct correlation between the internal comprehension of IT and the perceived complexity of cloud computing adoption and therefore the sourcing strategy. Despite stating that organisations need to create effective knowledge management strategies and processes to mitigate this, they do not investigate into what, if anything could be undertaken to improve this, or what effects the various decisions have on the public organisations. There are several barriers to technology adoption reported on throughout the dataset largely down to costs or organisations willingness to invest in R&D. More specifically, Acharya et al. (2022) acknowledged that financial constraints of public organisations impacting on decisions to invest in technology due to costs deprioritising technology agendas. What the report does not investigate however, is the risk in non-investment or action and what public organisations could do to bolster organisational capability in technology risk to support investment decisions.

Knowledge Management

As organisations must react to external factors such as policy changes, environmental developments and risks such as technological risks, it is imperative that they effectively manage knowledge and treat it as a strategic resource (Pacheco and Paul, 2023). Throughout the dataset, it is reported that organisations. Ravichandran (2005) states that to use or operate technology, firms must go further than theory and understand deeper, and practically how the technology works to diffuse the knowledge meaningfully within the organisation. This supports the ideas outlined in the above sections and promotes the idea of higher transaction costs when IT suppliers are involved. In this case, it is observed that IT suppliers may operate opportunistically and withhold information to initiate supplier lock-in (Clemons and Kleindorfer, 1992) or shirk responsibilities if they believe their client would be unable to detect or understand, as highlighted by Handley and Benton (2012).

Delving deeper into this issue, there are several barriers to knowledge management with regards to IT operations teams such as organisational barriers to understanding with examples being technical appreciation from a user perspective or business knowledge from a developer side (Al-Karaghouli et al., 2005). Other organisational factors include losing employees and capability when outsourcing the activity (Aron et al., 2005, Fauchart, 2006, Cheng and Chen, 2016). Other barriers to knowledge management may also be cost (Lean and Tucker, 2005, Meenan et al., 2010), or as outlined above, a supplier acting opportunistically.

To promote knowledge management, and improve the management of IT operations, organisations can use tools to manage knowledge, an example is the case study reported on by Meenan et al. (2010) whereby a bespoke wiki website was used to manage knowledge about supporting tools used in a radiology department. This tool was shown to be effective; however, the authors note that information is required to be constantly and consistently updated for it to be effective. This supports the above, however, the authors fail to report how this was implemented organisationally, or what the resource (staff or technical capability of workers) or financial overheads were to ensure success. Pistoni et al. (2022) propose that to combat suppliers behaving opportunistically and withholding information, specific contracts could be put in place to promote knowledge transfer and moderate the relationship. This does not however address the issue whereby the organisation would still be subject to unknown unknowns due to lower capability as reported in previous sections and therefore rely heavily on trust. It is not reported in the dataset how an organisation can optimally organise the internal IT operations department and what minimum level of technical capability is required to manage suppliers and reduce the risk of this type of opportunism.

Across the dataset and the above discussion in the other thematic areas, knowledge management emerges not only as a theme but as the key mitigation mechanism linking technological and relational risks. Specifically indicating that absorptive capacity and organisational capability could reduce vulnerability to opportunism by enabling organisations to detect shirking or information withholding.

Other practical knowledge management tools mechanisms are also shown to enable this capability such as codified knowledge (documentation, shared repositories, knowledge transfer) can directly weaken supplier power, enable knowledge retention to break the cycle as described above by preventing internal capability erosion over long-term relationships. This positions knowledge management as the central explanatory mechanism through which organisations navigate technological risk in outsourced IT operations, an insight not previously synthesised in the literature.

Conceptual Model

In totality, the above discussion indicates that risk in outsourced IT operations, which as described earlier is the operation and management ex-post, is not a set of isolated issues, but an interconnected system. Illustrated by a simple conceptual model, Figure 9, technological uncertainty increases information asymmetry which in turn increases the opportunity set for supplier opportunism.

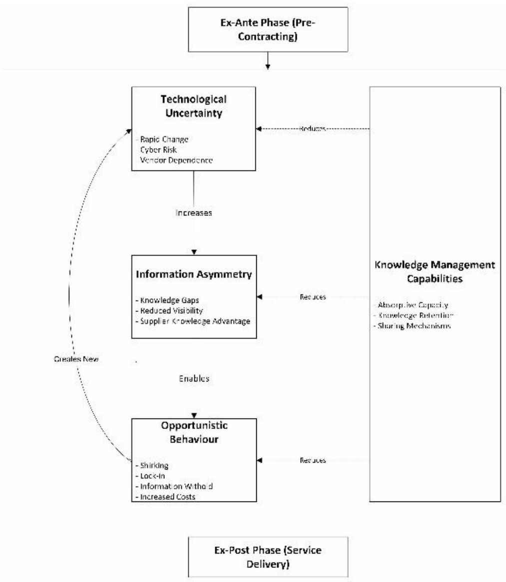

Figure 9 Knowledge management-centric risk mitigation in outsourced IT operations

This illustrates that the organisation's knowledge management capabilities moderate both effects. Where knowledge management is weak, organisations struggle to direct and monitor, but where knowledge management is strong, organisations can better interpret uncertainty, prioritise remediation and understand supplier behaviour.

RQ3: What are the Thematic gaps in the extant research for future research direction?

Upon addressing RQ1-2, there are several gaps which have been identified in the extant literature. Thematic gaps have been identified by illustrating Table 2 in

Figure 10

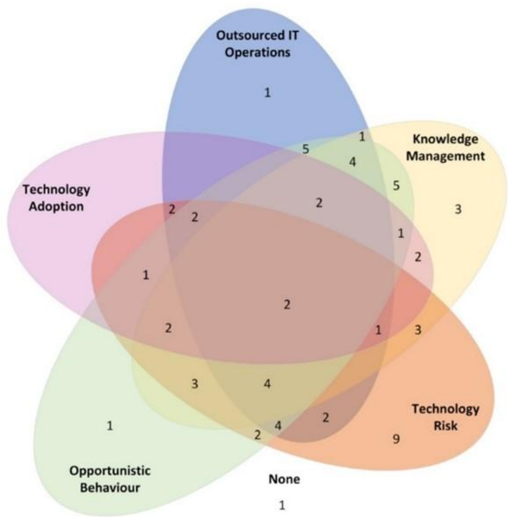

Figure 10 Venn diagram illustrating thematic split of dataset

Here, there are several thematic areas which have not been explored in isolation, these are:

1. Outsourced IT Operations + Technology Adoption
2. Outsourced IT Operations + Technology Adoption + Opportunistic Behaviour
3. Technology Adoption + Technology Risk
4. Technology Adoption + Technology Risk + Knowledge Management
5. Technology Adoption + Opportunistic Behaviour
6. Opportunistic Behaviour + Technology Risk
7. Opportunistic Behaviour + Knowledge Management
8. Outsourced IT Operations + Technology Risk + Knowledge Management

It is also worth noting that of the eight thematic gaps above, technology adoption appears five times, further indicating that the area of technology adoption is underreported in the extant literature in the field.

The discussion of the research themes in in response to RQ2 outline several thematic gaps which can be mapped directly to the above, these can be found in Table 5.

Gap in Research	**Thematic Gaps**
Effective management of IT suppliers ex-post and the maintenance required on the IT estate.	1
Using knowledge management and organisational design to mitigate technological risk and supplier lock-in.	8
Using Knowledge management to reduce supplier opportunistic behaviour	7
Using knowledge management to bolster organisational capability to make informed decisions on technology adoption to mitigate technology risk	4

Table 5 Matrix of research gaps to thematic gaps in research

Beyond the thematic combinations, the synthesis analysis highlights a deeper conceptual gap. Most studies focus on ex-ante IT outsourcing decisions, leaving the ex-post operation phase, where information asymmetry and opportunistic behaviour materialise largely unexplored, and although knowledge management is often mentioned as a mitigating factor, little research examples long-term knowledge retention or management strategies in connection with the relationship illustrated by Figure 9 despite its importance in limiting supplier dependence and detecting opportunistic behaviour. This is more deeply enhanced by public sector underreporting due to the longevity of IT operation contracts feeding into the cycle cadence outlined by Figure 9.

5.0. Future Research

Due to the low number of articles reporting primarily on public sector organisations within the field, there is an indication of a barrier to research here providing three possible gaps, the first being to research into how the public sector approaches IT operations when adopting an outsourcing model. Secondly, how this is comparative with the private sector, and finally more broadly, what barriers to research are preventing publication in this field.

As analysed in section 4, there are several thematic areas which researchers may wish to focus on. Any of the eight gaps in groupings of research themes could be a starting point, however, more specifically as an output of literature analysis, there is one main research gap which should be investigated further. The relationship between knowledge management, technological uncertainty, knowledge asymmetry and opportunistic behaviour during the ex-post phase of IT operations in the public sector.

Other areas to focus on as output from the analysis specifically centre around knowledge management within IT operations teams. How can these internal teams use knowledge management tools and techniques to mitigate uncertainties such as technological uncertainties, supplier lock-in and opportunistic behaviour, and how can this feed into technology adoption strategies and processes to ensure the organisation remains current and effective.

6.0. Conclusion

This paper presented a literature review, systematically analysing extant research in IT operations management when outsourcing the business process, specifically delving into the issues highlighted above. Findings reveal that as organisations continue to harness the power of technology in a world where technology development and innovation is increasing exponentially, they must continue to effectively manage their IT estates. Since the boom of IT outsourcing began in 1989, organisations have been enabled by IT supply organisations to utilise domain experts to innovate and run their IT estates to support and enhance their business processes with cost saving benefits.

This approach to managing IT does not come without its uncertainties however, with a trade-off in technical/specialist knowledge about the IT estate weaning with longer contracts and therefore the capability for effective decision making and technology uncertainty mitigation declining within the client organisation. Some argue that this is in the best interest of the supplier organisation and hence promotes opportunistic behaviour to drive higher transaction costs or supplier lock-in situations. Across the reviewed literature, knowledge management emerges as a critical mechanism to mitigate these risks, reducing information asymmetry and supporting more informed decision making both ex-ante and ex-post.

Gaps have been identified in the literature, specifically around knowledge management within the IT operations function with links out to supply chain management, technology risk management and technology adoption. More generally, research into the public sector is underreported giving researchers another lens to investigate current research themes with or adopt those previously listed. Notably, the ex-post phase of contracting, where most technological and relational risks materialise remains underreported, even more particularly in the public sector.

This paper has highlighted several challenges faced by IT operations and addressed these with mitigations. However, with the growth of the IT industry and therefore custom for IT supply chain organisations, that comes with an indication of continued growth and momentum in research within the field and expectations are this is set to increase over coming years to further enable organisations to operate optimally. As IT operations outsourcing continues to expand, future research should examine ex-post management, long term knowledge retention and sector specific differences will be essential for practitioners looking to strengthen organisational capability and mitigate uncertainty in outsourced IT operations.

References

Acharya, B., Lee, J. & Moon, H. (2022) 'Preference heterogeneity of local government for implementing ICT infrastructure and services through public-private partnership mechanism', *Socio-Economic Planning Sciences*, 79, 101103.

Adams, F.G. & Graham, K.W. (2017) 'Integration, knowledge creation and B2B governance: The role of resource hierarchies in financial performance', *Industrial Marketing Management*, 63, pp. 179–191.

Al-Karaghouli, W., Alshawi, S. & Fitzgerald, G. (2005) 'Promoting requirement identification quality', *Journal of Enterprise Information Management*, 18, pp. 256–267.

Al-Zubidy, A. & Carver, J.C. (2019) 'Identification and prioritization of SLR search tool requirements: An SLR and a survey', *Empirical Software Engineering*, 24, pp. 139–169.

Ali, O., Shrestha, A., Ghasemaghaei, M. & Beydoun, G. (2022) 'Assessment of complexity in cloud computing adoption: A case study of local governments in Australia', *Information Systems Frontiers*, 24, pp. 595–617.

Aria, M. & Cuccurullo, C. (2017) 'bibliometrix: An R-tool for comprehensive science mapping analysis', *Journal of Informetrics*, 11, pp. 959–975.

Aron, R., Clemons, E.K. & Reddi, S. (2005) 'Just right outsourcing: Understanding and managing risk', *Journal of Management Information Systems*, 22, pp. 37–55.

Asubiaro, T., Onaolapo, S. & Mills, D. (2024) 'Regional disparities in Web of Science and Scopus journal coverage', *Scientometrics*, 129, pp. 1469–1491.

Aubert, B.A., Houde, J-F., Patry, M. & Rivard, S. (2012) 'A multi-level investigation of information technology outsourcing', *The Journal of Strategic Information Systems*, 21, pp. 233–244.

Awe, O.A., Kulangara, N. & Henderson, D.F. (2018) 'Outsourcing and firm performance: A meta-analysis', *Journal of Strategy and Management*, 11, pp. 371–386.

Baho, S.A. & Abawajy, J. (2023) 'Analysis of consumer IoT device vulnerability quantification frameworks', *Electronics*, 12, 1176.

Biswas, B. & Mukhopadhyay, A. (2018) 'G-RAM framework for software risk assessment and mitigation strategies in organisations', *Journal of Enterprise Information Management*, 31, pp. 276–299.

Bui, Q.N., Leo, E. & Adelakun, O. (2019) 'Exploring complexity and contradiction in information technology outsourcing: A set-theoretical approach', *The Journal of Strategic Information Systems*, 28, pp. 330–355.

CABS (2021) *Academic Journal Guide 2021*. Available at: https://charteredabs.org/academic-journal-guide/academic-journal-guide-2021 (Accessed 15 May 2024).

Chang, Y.B., Gurbaxani, V. & Ravindran, K. (2017) 'Information technology outsourcing asset transfer and the role of contract', *MIS Quarterly*, 41, pp. 959–A3.

Cheng, J-H. & Chen, M-C. (2016) 'Influence of institutional and moral orientations on relational risk management in supply chains', *Journal of Purchasing and Supply Management*, 22, pp. 110–119.

Clemons, E.K. & Kleindorfer, P.R. (1992) 'An economic analysis of interorganizational information technology', *Decision Support Systems*, 8, pp. 431–446.

Cruz, J.M. & Liu, Z. (2011) 'Modeling and analysis of the multiperiod effects of social relationship on supply chain networks', *European Journal of Operational Research*, 214, pp. 39–52.

Duhadway, S., Talluri, S., Ho, W. & Buckhoff, T. (2022) 'Light in dark places: The hidden world of supply chain fraud', *IEEE Transactions on Engineering Management*, 69, pp. 874–887.

Durach, C.F., Kembro, J. & Wieland, A. (2017) 'A new paradigm for systematic literature reviews in supply chain management', *Journal of Supply Chain Management*, 53, pp. 67–85.

Eachempati, P. (2017) 'Change management in information asset', *Journal of Global Information Management*, 25, pp. 68–87.

Eze, S.C. et al. (2019) 'Determinants of perceived information need for emerging ICT adoption', *The Bottom Line*, 32, pp. 158–183.

Fauchart, E. (2006) 'Moral hazard and the role of users in learning from accidents', *Journal of Contingencies and Crisis Management*, 14, pp. 97–106.

Frigant, V. (2011) 'Are carmakers on the wrong track? Too much outsourcing in an imperfect-modular industry can be harmful', *International Journal of Manufacturing Technology and Management*, 22, pp. 324–343.

Hanafizadeh, P. & ZareRavasan, A. (2020) 'A systematic literature review on IT outsourcing decision and future research directions', *Journal of Global Information Management*, 28, pp. 160–201.

Handley, S.M. & Benton, W.C. (2012) 'The influence of exchange hazards and power on opportunism in outsourcing relationships', *Journal of Operations Management*, 30, pp. 55–68.

Hansen, C., Mena, C. & Aktas, E. (2019) 'The role of political risk in service offshoring entry mode decisions', *International Journal of Production Research*, 57, pp. 4244–4260.

Hansen, C., Mena, C. & Skipworth, H. (2017) 'Exploring political risk in offshoring engagements', *International Journal of Production Research*, 55, pp. 2051–2067.

Haq, S.U., Gu, D., Liang, C. & Abdullah, I. (2019) 'Project governance mechanisms and the performance of software development projects: Moderating role of requirements risk', *International Journal of Project Management*, 37, pp. 533–548.

Hoffmann, P., Schiele, H. & Krabbendam, K. (2013) 'Uncertainty, supply risk management and their impact on performance', *Journal of Purchasing and Supply Management*, 19, pp. 199–211.

Hong, E., Son, B-G. & Menachof, D. (2010) 'Exploring the link between IT systems and the outsourcing of logistics activities: A transaction cost perspective', *International Journal of Logistics Research and Applications*, 13, pp. 41–57.

Hoseini, E., van Veen, P., Bosch-Rekveldt, M. & Hertogh, M. (2020) 'Cost performance and cost contingency during project execution: Comparing client and contractor perspectives', *Journal of Management in Engineering*, 36, 05020006.

Hyde, K.F. (2000) 'Recognising deductive processes in qualitative research', *Qualitative Market Research*, 3, pp. 82–90.

Jacobs, J., Romanosky, S., Adjerid, I. & Baker, W. (2020) 'Improving vulnerability remediation through better exploit prediction', *Journal of Cybersecurity*, 6.

Kauffman, R.J. & Tsai, J.Y. (2009) 'The unified procurement strategy for enterprise software: A test of the "move to the middle" hypothesis', *Journal of Management Information Systems*, 26, pp. 177–204.

Khan, I., Rutherford, B.N. & Williams, A.J. (2019) 'Information technology outsourcing: Influence of supplier firm size and reputation on buyers' perceptions of opportunism and uncertainty', *Operations and Supply Chain Management*.

Kloyer, M. & Scholderer, J. (2012) 'Effective incomplete contracts and milestones in market-distant R&D collaboration', *Research Policy*, 41, pp. 346–357.

Lacity, M. & Willcocks, L. (2014) 'Business process outsourcing and dynamic innovation', *Strategic Outsourcing*, 7, pp. 66–92.

Lean, J. & Tucker, J. (2005) 'Evolving a regionally-based mechanism for the provision of technical knowledge to SMEs', *Industry and Higher Education*, 19, pp. 325–332.

Leverett, É., Rhode, M. & Wedgbury, A. (2022) 'Vulnerability forecasting: Theory and practice', *Digital Threats*, 3, Article 42.

Lin, J., Zhang, H., Adams, B. & Hassan, A.E. (2023) 'Vulnerability management in Linux distributions', *Empirical Software Engineering*, 28, 47.

Lin, T. & Hekkala, R. (2016) 'Governance structure in IT outsourcing: A network perspective', *Strategic Outsourcing*, 9, pp. 38–59.

Lin, T. & Vaia, G. (2015) 'The concept of governance in IT outsourcing: A literature review'.

Lowis, L. & Accorsi, R. (2011) 'Vulnerability analysis in SOA-based business processes', *IEEE Transactions on Services Computing*, 4, pp. 230–242.

Mantha, B.R.K. & García de Soto, B. (2021) 'Assessment of the cybersecurity vulnerability of construction networks', *Engineering, Construction and Architectural Management*, 28, pp. 3078–3105.

Marinč, M. (2013) 'Banks and information technology: Marketability vs. relationships', *Electronic Commerce Research*, 13, pp. 71–101.

Mathew, S.K. & Chen, Y. (2013) 'Achieving offshore software development success: An empirical analysis of risk mitigation through relational norms', *The Journal of Strategic Information Systems*, 22, pp. 298–314.

Meenan, C. et al. (2010) 'Use of a wiki as a radiology departmental knowledge management system', *Journal of Digital Imaging*, 23, pp. 142–151.

Mishra, R., Singh, R.K. & Papadopoulos, T. (2024) 'Linking digital orientation and data-driven innovations: A SAP–LAP linkage framework and research propositions', *IEEE Transactions on Engineering Management*, 71, pp. 1346–1358.

Moher, D. et al. (2009) 'Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement', *BMJ*, 339, b2535.

Nduwimfura, P. & Zheng, J. (2015) 'A review of risk management for information systems outsourcing', *International Journal of Business, Humanities and Technology*, 5, pp. 28–33.

Nepomuceno, T.C.C. et al. (2022) 'When penalty fails: Modelling contractual misincentives with evidence from Portugal ITO agreements', *Sage Open*, 12, 21582440221141850.

O'Neill, M., Booth, S. & Lamb, J. (2018) 'Using NVivo for literature reviews: The eight step pedagogy (N7+1)', *The Qualitative Report*, 23, pp. 21–39.

Öbrand, L., Augustsson, N-P., Mathiassen, L. & Holmström, J. (2019) 'The interstitiality of IT risk: An inquiry into information systems development practices', *Information Systems Journal*, 29, pp. 97–118.

Pacheco, C. & Paul, B. (2023) 'Applying complexity theory perspective to knowledge management in the innovation context', *VINE Journal of Information and Knowledge Management Systems*, ahead-of-print.

Palomino, M.A., Vincenti, A. & Owen, R. (2013) 'Optimising web-based information retrieval methods for horizon scanning', *Foresight*, 15, pp. 159–176.

Pearce, J.M. (2018) 'How to perform a literature review with free and open-source software', *Practical Assessment, Research & Evaluation*, 23.

Pilbeam, C., Alvarez, G. & Wilson, H. (2012) 'The governance of supply networks: A systematic literature review', *Supply Chain Management*, 17, pp. 358–376.

Pistoni, A., Arcari, A. & Gigliarano, C. (2022) 'Managerial control systems and innovation partnership success: An empirical analysis in Italian firms', *European Journal of Innovation Management*, ahead-of-print.

Qin, X. et al. (2018) 'Analysing manufacturer and insurance-based risk mitigation policy with equipment service contracting', *Enterprise Information Systems*, 12, pp. 1359–1381.

Qu, W.G. & Pinsonneault, A. (2011) 'Country environments and the adoption of IT outsourcing', *Journal of Global Information Management*, 19, pp. 30–50.

Ravichandran, T. (2005) 'Organizational assimilation of complex technologies: An empirical study of component-based software development', *IEEE Transactions on Engineering Management*, 52, pp. 249–268.

Refaiy, M. & Labib, A. (2009) 'The effect of applying tacit knowledge on maintenance performance', *Knowledge Management Research & Practice*, 7, pp. 277–288.

Rustagi, S., King, W.R. & Kirsch, L.J. (2008) 'Predictors of formal control usage in IT outsourcing partnerships', *Information Systems Research*, 19, pp. 126–143.

Sabherwal, R. & King, W.R. (1991) 'Towards a theory of strategic use of information resources: An inductive approach', *Information & Management*, 20, pp. 191–212.

Samtani, S., Kantarcioglu, M. & Chen, H. (2020) 'Trailblazing the artificial intelligence for cybersecurity discipline: A multi-disciplinary research roadmap', *ACM Transactions on Management Information Systems*, 11, Article 17.

Sauer, P.C. & Seuring, S. (2023) 'How to conduct systematic literature reviews in management research: A guide in 6 steps and 14 decisions', *Review of Managerial Science*, 17, pp. 1899–1933.

Schneider, S. & Sunyaev, A. (2016) 'Determinant factors of cloud-sourcing decisions: Reflecting on the IT outsourcing literature in the era of cloud computing', *Journal of Information Technology*, 31, pp. 1–31.

Silic, M. & Back, A. (2016) 'The influence of risk factors in decision-making process for open-source software adoption', *International Journal of Information Technology & Decision Making*, 15, pp. 151–185.

SJR (2024) *Journal Rankings*. Available at: https://www.scimagojr.com/journalrank.php (scimagojr.com in Bing) (Accessed 15 May 2024).

Syed, R. (2020) 'Cybersecurity vulnerability management: A conceptual ontology and cyber intelligence alert system', *Information & Management*, 57, 103334.

Szczepański, R. & Światowiec-Szczepańska, J. (2012) 'Risk management system in business relationships—Polish case studies', *Industrial Marketing Management*, 41, pp. 790–799.

Tang, M., Alazab, M., Luo, Y. & Donlon, M. (2018) 'Disclosure of cyber security vulnerabilities: Time series modelling', *International Journal of Electronic Security and Digital Forensics*, 10, pp. 255–275.

Valiente, M-C., Garcia-Barriocanal, E. & Sicilia, M-A. (2012) 'Applying an ontology approach to IT service management for business-IT integration', *Knowledge-Based Systems*, 28, pp. 76–87.

Varpio, L., Paradis, E., Uijtdehaage, S. & Young, M. (2020) 'The distinctions between theory, theoretical framework, and conceptual framework', *Academic Medicine*, 95.

Wilkin, C.L., Couchman, P.K., Sohal, A. & Zutshi, A. (2016) 'Exploring differences between smaller and large organizations' corporate governance of information technology', *International Journal of Accounting Information Systems*, 22, pp. 6–25.

Worrell, J.L., Di Gangi, P.M. & Bush, A.A. (2013) 'Exploring the use of the Delphi method in accounting information systems research', *International Journal of Accounting Information Systems*, 14, pp. 193–208.

Yang, N., Guo, M., Wang, J. & Zhang, Y. (2021) 'The moderating effect of network power on relational risks and knowledge flow in R&D network', *Management Decision*, 59, pp. 2421–2441.

Yang, N., Song, Y., Zhang, Y. & Wang, J. (2020) 'Dark side of joint R&D collaborations: Dependence asymmetry and opportunism', *Journal of Business & Industrial Marketing*, 35, pp. 741–755.

Yigitbasioglu, O., Mackenzie, K. & Low, R. (2013) 'Cloud computing: How does it differ from IT outsourcing and what are the implications for practice and research?', *International Journal of Digital Accounting Research*, 13, pp. 99–121.

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 75-85
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1807

The Effect of Mobile Payment Methods on Customer Decisions on Jumia's Shopping Platform in Nigeria

[1]*Promise Akwaowo*
[2]*Imani Silver Kyaruzi*
[1]*QAHE- Ulster University (Graduate)*
[2]*QAHE- Ulster University*

Abstract
This paper analyses how mobile payment technologies affect customer buying behaviour on Jumia, a major e-commerce site in Nigeria. Using a quantitative survey of 150 respondents, the study examines three main areas: (1) whether mobile payment options improve conversion rates, (2) whether they build customer trust, and (3) how infrastructure issues like internet and smartphone access play a role. The results show that customer trust strongly predicts both mobile payment adoption and actual purchases ($r = 0.451, p < 0.001$), whereas infrastructure challenges have a minimal direct impact ($r = 0.018, p = 0.811$). Correlation analysis further confirms a moderate, positive link between trust and conversion ($r = 0.456, p < 0.01$). These findings highlight trust and perceived-security as essential for successful digital transactions. The implication is that, although infrastructure issues remain, improved platform reliability is making them less important. The research offers business management insights for building digital trust and simplifying payment systems as mobile commerce gains ground in Nigeria.

Keywords: Mobile payments, Customer trust, Jumia, E-commerce, Nigeria.

Wordcount: 172

1.0 Introduction

Waves of digital transformation have fundamentally changed consumer behaviour and corporate strategy, restructuring the global retail market (Gomber et al., 2018; Lottu et al., 2023). While this shift has been gradual in developed economies, it has been highly disruptive in developing countries like Nigeria. It requires reevaluating long-standing assumptions about consumer decision-making and platform competitiveness. Rapid growth in artificial intelligence, big data analytics, and online payments has altered consumer information searches, transaction processes, and decision-making, which previously relied on traditional factors like price and product quality (Khrais, 2020). Digital consumers increasingly seek convenience, transaction speed, personalisation, and most importantly, the perceived reliability of payment systems, which affects purchase completion and retention (Islam, 2024). Development gaps and socio-economic constraints complicate e-commerce in Nigeria, making consumer insights a strategic necessity. Data-driven e-commerce systems now actively shape buying outcomes but also face growing scrutiny regarding payment security and privacy. Jumia, a leading Nigerian e-commerce platform, uses advanced tech to boost customisation and marketing. However, improved technology also raises concerns about payment security, underscoring the importance of trust in reliable payment systems for user conversion.
Mobile payments now drive online trade in sub-Saharan Africa, especially in areas with limited banking access. In Nigeria, fintech solutions mobile wallets like OPay, PalmPay, Paga, and gateways like Flutterwave and Paystack enable secure, flexible, real-time transactions. Their integration into platforms such as Jumia has reduced the use of cash on delivery, which was historically prone to high returns, logistics issues, and security concerns (Anjum & Chai, 2020). Functionally, mobile payments lower checkout friction and ease transactions, likely improving conversion rates.

Despite efficiency gains, mobile payments present cybersecurity risks and service issues for Nigerian users. Persistent transaction losses and unclear dispute processes undermine trust, creating a paradox of adoption and scepticism. Mobile payment adoption in developed economies occurs within regulated, well-supported environments, yielding less friction and greater trust (Putrevu & Mertzanis, 2024). In contrast, Nigeria's mobile payment environment is shaped by socio-technical factors influencing the link between availability and usage (Dhar et al., 2025). On platforms such as Jumia, simply offering mobile payment does not ensure use or conversion. Demographic and contextual variables—including income, digital literacy, location, product risk, and transaction value further shape consumer behaviour (Lindheim & Grimsrud, 2017). This underscores the need to analyse how mobile systems, trust, infrastructure, and diverse consumers interact to drive conversion in new e-commerce markets. Instead of assuming linear adoption, this study examines mobile payments within a complex nexus of technology, behaviour, and constraints (Anagreh et al., 2024). Digital technologies are spreading in Nigerian retail, yet little research exists on how mobile payment systems affect consumer behaviour at conversion on major e-commerce platforms. Existing studies mainly address fintech adoption or use, rather than transaction completion in specific platform ecosystems. Jumia, a prominent African e-commerce platform, provides a context for exploring the relationship between mobile financial technologies and consumer decision-making in emerging markets. Electronics lead Nigeria's e-commerce sales, with user growth expected to reach nearly 40% by 2029. Understanding how mobile payment systems shape customer confidence and purchase decisions is essential as the market grows.

Despite the increased availability of advanced mobile payment services, transactional friction still defines consumer experiences on websites like Jumia. This paper addresses the conflict between accessible technology and user trust. While digital wallets and mobile money systems are marketed as fast and convenient, studies show that transaction failures, slow refunds, and perceived security breaches increase mistrust and can reduce completed purchases (Ankrah et al., 2024). For high-value purchases such as electronics, Nigerian customers often prefer cash-on-delivery when mobile payments seem unreliable or opaque (Ezuwore-Obodoekwe et al., 2014). This shift presents a strategic challenge for e-commerce businesses aiming to reduce operational costs and cash-handling risks while maintaining conversion rates (Badran, 2021). To fill this gap, the research initially examines the impact of various mobile payment services on customer conversion to buyers on Jumia, compared with conventional payment methods such as cash-on-delivery. This research question builds on the concept of broad adoption narratives, specifically examining conversion behaviour in a platform-specific environment and thereby adding quantitative data to an under-reported aspect of mobile commerce research.

Nevertheless, it cannot be assumed that technological functionality alone is sufficient to produce the observed behavioural variation. Trust is another important predictor of digital transaction behaviour that has been gradually acknowledged in the existing e-commerce literature. Still, its mediating nature is under-researched in the context of e-commerce in Nigeria (Ghali, 2024). This study, therefore, examines the extent to which customer trust affects the adoption of mobile payment options on Jumia and how this relationship differs across demographic groups by age and income. By serving as a mediating role between payment technology and conversion results, the study will fill a significant gap in the literature and connect the heterogeneity of consumer reactions.

Finally, the study acknowledges that Nigeria's structural and infrastructural constraints—including inconsistent broadband connectivity, uneven smartphone penetration, and regulatory ambiguity continue to shape digital payment experiences (Nucciarelli & Sadowski, 2018). Instead of viewing infrastructure as an unalterable constraint, the study critically examines the hypothesis that these constraints indirectly affect the use of mobile payments by shaping trust, thereby challenging techno-optimistic beliefs prevalent in much of the fintech literature. By combining these research questions into a sceptical, context-specific framework, the research breaks the linear model sequence of fintech diffusion. It redefines mobile payments as a facilitator and a possible bottleneck to consumer conversion. By doing so, it contributes to research on digital commerce in new markets. It provides practical information to platforms like Jumia as they strive to balance operational efficiency with long-term consumer trust.

Research questions

1. How do various mobile payment methods (e.g., digital wallets, mobile money) affect customer conversion rates on Jumia's platform compared to traditional payment options such as cash on delivery?
2. What is the impact of customer trust on Jumia's adoption of mobile payment methods, and how does this trust affect conversion rates across different demographic segments (e.g., age and income level)?
3. How do infrastructural constraints (internet connectivity and smartphone access) influence customer trust and indirectly affect mobile payment–driven conversion behaviour on Jumia?

2.0 Theoretical Framework and Hypotheses

This paper is anchored in an integrative theoretical framework based on Customer Trust Theory, the Technology Acceptance Model (TAM), and the Theory of Planned Behaviour (TPB). A combination of these frameworks allows for explaining the impact of mobile payment systems on customer conversion behaviour on the Jumia e-commerce platform in the Nigerian context, where technological adoption is combined with institutional uncertainty and infrastructural limitations.

*The Customer Trust Theory***,** as explained by Roger Mayer, James Davis, and David Schoorman (1995), states that trust is a perception of ability, benevolence, and integrity. Trust minimises perceived transactional risk and uncertainty in digital commerce settings, especially when dealing with financial information and irreversible payments (Islam, 2024). In mobile payment ecosystems, perceived security, system reliability, and transparency are vital antecedents of trust. This theory becomes particularly relevant in emerging markets, such as Nigeria, where consumer risk sensitivity is heightened by prior experiences of unsuccessful transactions, delayed refunds, and fraud (Ifechukwu, 2022). In this regard, the concept of trust is based on the study of Simatele. (2024), It is not approached as a background condition in the research, but rather as a key mediating process through which mobile payment technologies are translated into actual purchase success.

The Technology Acceptance Model (TAM), introduced by Fred Davis (1989), further supports the framework by elucidating the use of technology in terms of perceived usefulness and perceived ease of use (Alturki & Aldraiweesh, 2022). According to TAM, consumers are more likely to embrace a technology that improves task performance and requires minimal effort. Mobile payment systems can be viewed as beneficial for Jumia, as they are quicker to check out and less uncomfortable than cash-on-delivery. Nevertheless, TAM has been criticised for lacking adequate explanatory power in high-risk environments, particularly regarding trust and contextual constraints (Burgess et al., 2023). This research thus builds on TAM by incorporating customer trust as a mediating variable, as perceived usefulness alone cannot necessarily lead to conversion when trust in payment security is low.

According to Icek Ajzen (1991), the *Theory of Planned Behaviour (TPB)* is an addition to the framework that includes perceived behavioural control, which is especially applicable in infrastructurally constrained settings (Hagger et al. 2022). TPB clarifies that attitudes do not solely define behavioural intention; perceived control over external conditions also influences it. In Nigeria, infrastructure issues, such as the reliability of internet connectivity and smartphone access, affect consumers' confidence in completing mobile payment transactions successfully (Famosaya, 2024). Nevertheless, TPB admits the possibility that this sort of constraint could have an indirect rather than a direct impact, particularly where users have adopted adaptive behaviours or are supported by platform-level optimisation.

Combining these theories, the research paper views mobile payment availability and infrastructure quality as independent variables, customer trust as a mediating variable, and customer conversion rate as a dependent variable. This comprehensive framework goes beyond binary adoption frameworks by describing the process and the reasons why mobile payment systems affect conversion behaviour in a state of partial infrastructural sufficiency and changing digital maturity. Based on the theoretical framework and the stated research questions, the following hypotheses are formulated to guide the empirical investigation:

H1: Customer trust in mobile payment systems has a statistically significant and positive effect on customer conversion rates on Jumia's shopping platform, such that higher perceived security and reliability increase the likelihood of purchase completion.
H2: Infrastructure challenges, including limited internet connectivity and restricted smartphone access, have a statistically significant negative effect on customer trust in Jumia's mobile payment systems, thereby increasing perceived transaction risk.
H3: Infrastructure challenges do not exert a statistically significant direct effect on customer conversion rates on Jumia; rather, their influence on conversion behaviour operates indirectly through their impact on customer trust.
Collectively, these hypotheses facilitate a rigorous examination of whether customer trust functions as the principal explanatory mechanism linking mobile payment systems to conversion behaviour. By distinguishing between direct and indirect effects, the hypotheses advance theoretical understanding of the dynamics of digital commerce in emerging markets, where infrastructural constraints coexist with growing digital payment adoption.

3.0 Methodology

3.1. Research Design

The research design was a quantitative, cross-sectional survey that followed a positivist, deductive research philosophy, as described by the Research Onion by Saunders, Lewis, and Thornhill (2019). The positivist position was suitable because the study aimed to test a hypothesis developed from a theory with measurable variables and statistical analysis (Ali, 2024). It was concluded that a cross-sectional design is the most appropriate method for assessing respondents' perceptions of the mobile payment system, trust, and infrastructural conditions at a single point in time, consistent with previous empirical research in digital commerce and technology adoption studies.

3.2. Sampling Strategy and Data Collection.

Following the methodological advice of Saunders et al. (2019) for access-constrained populations, non-probability convenience sampling was used. The target population included Nigerian Jumia users with prior experience with online shopping and mobile payment methods. The recruitment of respondents was based on online platforms, such as social networks and online consumer discussions, which ensured applicability to the research environment. One hundred and fifty (150) valid returns were given. This is sufficient to perform a minimum analysis of multiple regression as it has more than the recommended number of observations per predictor (10-15) (Pate et al., 2023). Although convenience sampling hinders statistical generalisability, it is unanimously adopted in more exploratory and platform-specific e-commerce studies, especially when the sampling frame is unavailable.

3.3. Measurement Instrument

A structured questionnaire implemented closed-ended questions of a five-point Likert scale between Strongly Disagree (1) and Strongly Agree (5) was used to collect data. The measurement items were based on questions from well-known scales used in previous research on mobile payment adoption, customer trust, and digital transaction behaviour, and they possess content validity. **The questionnaire captured four core constructs:** Mobile payment availability, Infrastructure challenges, Customer trust and Customer conversion behaviour. All items were contextualised to Jumia's platform to enhance construct relevance and respondent comprehension.

Reliability and Validity Assessment

Internal consistency reliability was assessed using Cronbach's alpha coefficient. The results exceeded the recommended 0.70 threshold, confirming acceptable reliability across constructs: Customer Trust (α = 0.81), Conversion Behaviour (α = 0.78) and Infrastructure Challenges (α = 0.74). These values indicate strong internal consistency and support the instrument's suitability for hypothesis testing. Construct validity was further supported through alignment with established theoretical frameworks (Customer Trust Theory, TAM, and TPB), consistent with Saunders et al.'s (2019) emphasis on theoretical coherence.

Variable Specification and Model Construction

The study operationalised variables as follows:
Independent Variables (IV): Customer Trust (CT)(Mediator); Infrastructure Challenges (IC)
Dependent Variable (DV): Customer Conversion Rate (CR)
The regression model was specified as: $CR=\beta0+\beta1(CT)+\beta2(IC)+\varepsilon$
Where:
CR represents customer conversion behaviour,
CT denotes customer trust in mobile payment systems,
IC reflects perceived infrastructure challenges,
$\beta0$ is the constant,
$\beta1$ and $\beta2$ are regression coefficients,
E is the error term.
This specification enabled direct testing of the hypothesised relationships and the relative explanatory power of trust versus infrastructure.

3.4. Data Analysis Techniques

Data analysis was conducted using IBM SPSS Statistics (Version 2023). Analytical procedures included: Descriptive statistics to summarise respondent perceptions, Pearson correlation analysis to examine bivariate relationships, Multiple regression analysis to test hypotheses and assess predictive effects. This analytical strategy aligns with Saunders et al.'s (2019) recommendation for methodological consistency between research questions, hypotheses, and statistical techniques.

4.0 Findings

This uncovers data-driven insights into how mobile payment innovations influence customer behaviour and purchasing decisions across Jumia's Nigerian e-commerce platform.

Table 1 "Research Question 1: How do various mobile payment methods (e.g., digital wallets, mobile money) affect customer conversion rates on Jumia's platform compared to traditional payment options such as cash on delivery?"

Items	Questions	SA	A	N	D	SD	Total
1	I find mobile payment methods more convenient than traditional payment options.	68 (45.3%)	54 (36.0%)	9 (6.0%)	12 (8.0%)	7 (4.7%)	150 (100%)
2	Using mobile payment methods increases my likelihood of completing a purchase on Jumia.	42 (28.0%)	81 (54.0%)	14 (9.3%)	10 (6.7%)	3 (2.0%)	150 (100%)
3	I trust mobile payment methods more than traditional payment options like cash on delivery.	18 (12.0%)	43 (28.7%)	45 (30.0%)	35 (23.3%)	9 (6.0%)	150 (100%)
4	Mobile payment methods save me time during the checkout process.	39 (26.0%)	85 (56.7%)	15 (10.0%)	6 (4.0%)	5 (3.3%)	150 (100%)
5	I prefer shopping on platforms that offer mobile payment methods over those that do not.	36 (24.0%)	58 (38.7%)	35 (23.3%)	16 (10.7%)	5 (3.3%)	150 (100%)

Based on the findings in this study, conversion rates on Jumia compared to traditional methods, such as cash-on-delivery. The importance of convenience is evident as the chief factor, as more than 81 per cent of participants reported that mobile payments make the shopping process easier and predispose them to making more purchases. As a matter of fact, 54 per cent concur, and 28 per cent strongly agree that mobile payment options have a direct positive effect on purchase decisions.

Time efficiency is also a notable attribute, with 56.7% indicating that mobile payments facilitate a smooth checkout. Friction and cart abandonment are important metrics in improving the performance of the entire platform. Nevertheless, trust is an evolving factor: although convenience drives adoption, only 12% trust mobile payments over cash-on-delivery, indicating that security should be enhanced and communication with customers improved to accelerate adoption.

Moreover, 38.7% show a clear preference for platforms offering mobile payment options, reflecting a growing shift toward digital-first shopping experiences. Overall, with 27.06% strongly agreeing and 42.62% agreeing that mobile payments improve their shopping experience, the data underscores their role in enhancing customer engagement, optimising checkout flows, and driving higher conversion rates across Jumia's e-commerce ecosystem.

Table 2: "Research Question 2: What is the impact of customer trust on Jumia's adoption of mobile payment methods, and how does this trust affect conversion rates across different demographic segments (e.g., age and income level)?"

Items	Questions	SA	A	N	D	SD	Total
1	I trust Jumia's mobile payment methods to protect my financial information.	25 (16.7%)	72 (48.0%)	39 (26.0%)	8 (5.3%)	6 (4.0%)	150 (100%)
2	Trust in Jumia's mobile payment methods increases my likelihood of making purchases on the platform.	23 (15.7%)	84 (56.0%)	29 (19.3%)	10 (6.7%)	4 (2.7%)	150 (100%)
3	The security features of Jumia's mobile payment methods positively influence my decision to use them over traditional payment options.	19 (12.7%)	75 (50.0%)	43 (28.7%)	10 (6.7%)	3 (2.0%)	150 (100%)
4	I am more likely to use Jumia's mobile payment methods if I see positive	42 (28.0%)	78 (52.0%)	17 (11.3%)	8 (5.3%)	5 (3.3%)	150 (100%)

	reviews and testimonials from other users.						
5	My trust in Jumia's mobile payment methods is higher if the platform offers transparent information about security measures and data protection.	45 (30.0%)	78 (52.0%)	16 (10.7%)	5 (3.3%)	6 (4.0%)	150 (100%)

The survey also shows that the relationship between customer trust and the use of Jumia mobile payment systems is strong, and it directly influences conversion rates across demographics. Financial security is a significant motivational factor, as 64.7% of respondents believe Jumia can protect their payment information, which would be a defining factor in their ability to embrace digital payments. The transfer of trust into purchasing desire likewise yields 71.7%, indicating that safe payment systems make them more inclined to make purchases on the platform. The most significant finding was that 62.7 per cent confirmed that they influence the decision to use mobile payments rather than traditional payments. Positive user reviews complement it: 80 per cent report that testimonials increase their confidence in mobile payment systems. In addition, 82 per cent highlight the relevance of open communication regarding data security, which is supported by the fact that trust, which is supported by effective security and effective communication, is important in fostering mobile payments and conversion rates at Jumia.

Table 3: Research Question 3: How do infrastructure challenges, including limited internet connectivity and smartphone penetration in certain regions of Nigeria, affect?

Items	Questions	SA	A	N	D	SD	Total
1	Limited internet connectivity in my area makes it difficult to use mobile payment methods on Jumia.	22 (14.0%)	54 (36.0%)	28 (18.7%)	32 (21.3%)	14 (9.0%)	150 (100%)
2	The availability of reliable internet connectivity positively influences my decision to use mobile payment methods on Jumia.	30 (20.0%)	74 (49.3%)	27 (18.0%)	9 (6.0%)	10 (6.7%)	150 (100%)
3	Limited smartphone penetration in my region affects my ability to use mobile payment methods on Jumia.	16 (10.7%)	37 (24.7%)	30 (20.0%)	51 (34.0%)	16 (10.7%)	150 (100%)
4	I am more likely to use mobile payment methods on Jumia if I have access to a smartphone with internet connectivity.	55 (36.7%)	65 (43.3%)	20 (13.3%)	7 (4.7%)	3 (2.0%)	150 (100%)
5	Infrastructure challenges, such as poor internet connectivity and lack of smartphones, negatively impact my shopping experience in Jumia.	35 (23.3%)	70 (46.7%)	20 (13.3%)	14 (9.3%)	11 (7.3%)	150 (100%)

The survey examines the impact of infrastructure issues on the acceptance of mobile payments on the Jumia platform in Nigeria. The major challenges include poor internet connectivity, with half of the respondents (50.7) saying that these services are a constraint on the effective utilisation of mobile payments; hence, there is a need for a good internet connection to make transactions effective. Conversely, 69.3 of them indicate that the availability of a stable internet connection increases the likelihood of moving to mobile payment systems with a significant impact, and that there is a direct relationship between infrastructural quality and adoption rates.

Smartphone penetration also plays a big role: 35.4% acknowledge that a lack of smartphone access is a restraining factor in utilising mobile payments, and 80% are sure they have a smartphone and that internet access is a key factor in their likelihood of utilising digital payment options. Generally, about 70 per cent of the respondents believe that poor connectivity and the unavailability of smartphones negatively affect their online shopping experience. These findings indicate that the digital infrastructure should be enhanced to enhance the user experience and encourage the utilisation of mobile payments on the Jumia platform.

Table 4: Regression

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.456[a]	.208	.197	3.10852

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	372.885	2	186.442	19.295	.000[b]
Residual	1420.448	147	9.663		
Total	**1793.333**	**149**			

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	5.109	1.215		4.204	.000
Customer trust	.523	.088	.451	5.967	.000
Infrastructure challenges	.019	.078	.018	.240	.811

Source: IBM SPSS 2023

The regression model will quantify the influence of customer trust and infrastructure issues on the decision to implement mobile payments on the Jumia platform. The results show a moderately strong correlation (R = 0.456) with 20.8 per cent of the variance (R 2 = 0.208), indicating the model's significance (F = 19.295, p = 0.001).

The strongest predictor is customer trust, with a standardised Beta of 0.451 ($p < 0.001$), indicating a positive and significant impact on the effectiveness of mobile payments. It is also consistent with Customer Trust Theory and the Technology Acceptance Model, which focus on the role of trust in reducing perceived risk and enhancing digital adoption.

On the other hand, infrastructure issues are not much of a concern (Beta = 0.018, p = 0.811), suggesting that Jumia users are more worried about security and reliability than connection issues. This is the opposite of the former research, which found that infrastructure was a significant barrier, suggesting greater digital maturity and the efficient design of user-friendly payment systems that transcend Jumia's infrastructure limitations.

Table 5: Correlation
Relationship among Conversion Rate, Customer Trust, and Infrastructure Challenges

Variables	Conversion Rate	Customer Trust	Infrastructure Challenges
Conversion Rate	—	.456**	.127
Customer Trust	.456**	—	.241**
Infrastructure Challenges	.127	.241**	—

Source: IBM SPSS 2023

In analysing the impact of mobile payment methods on customer decisions within Jumia's Nigerian platform, the correlation and hypothesis tests provide key insights.
Correlation Analysis:
Conversion Rate & Customer Trust: A Pearson correlation coefficient of r = 0.456 ($p < 0.01$) indicates a moderate-to-strong, statistically significant relationship, suggesting that higher customer trust is associated with higher conversion rates.
Conversion Rate & Infrastructure Challenges: With r = 0.127 (p = 0.122), the relationship is weak and insignificant, suggesting that infrastructure issues have a minimal direct effect on conversion rates.
Customer Trust & Infrastructure Challenges: A moderate negative correlation (r = 0.241, $p < 0.01$) reveals that infrastructure issues erode customer trust, indirectly influencing conversions.

5.0. Discussion and Implications for Jumia and Digital Technology

The results of this research present detailed insights into how mobile payment systems affect customer conversion behaviour on the Jumia platform. Specifically, the findings indicate that customer trust is the most dominant explanatory variable, whereas infrastructural challenges have a more indirect effect on conversion outcomes. This difference is important when interpreting mobile payment adoption and usage in emerging e-commerce markets like Nigeria.

Customer trust was found to be a statistically significant, positive predictor of conversion behaviour, indicating that it plays a central role in purchase decisions facilitated by mobile payments. This observation aligns with the Customer Trust Theory, which holds that trust reduces perceived transactional risk and uncertainty, especially in settings involving financial data exchange (Budiharseno and Kim, 2023). Within the Jumia framework, perceptions of the security, reliability, and transparency of mobile payment systems boost users' confidence, thereby improving the likelihood of completing transactions. This finding is also consistent with the Technology Acceptance Model (TAM), which postulates that perceived usefulness and ease of use will encourage the acceptance of technology when its users have confidence in system reliability (Kamal et al., 2020). Empirically, the research supports previous findings that attributes of trust, including platform security assurances and platform credibility, are determinants of online payment behaviour (Albshaier et al., 2024).

Conversely, infrastructural issues, such as low internet connectivity and low smartphone coverage, were reported to have no statistically significant direct impact on conversion behaviour. This finding contrasts with previous research, which names infrastructure as one of the greatest obstacles to the uptake of digital payments in developing economies. Yet, rather than conflicting with the current literature, the findings indicate a shift in user behaviour towards a more structural approach in platform-mediated settings such as Jumia. In particular, users admit that infrastructural restrictions do not affect their ability to make a purchase directly, given the high quality of the payment systems used on the platform.

Notably, the correlation analysis shows that infrastructural challenges are strongly correlated with low customer trust, suggesting an indirect influence. In this respect, infrastructure has a conversion behaviour influenced by trust, not a transactional constraint. This mediating association helps address the seeming contradiction between respondents' perceptions of the infrastructural challenges and the negligible regression coefficient for infrastructure. It is also compatible with the Theory of Planned Behaviour (TPB), which highlights perceived behavioural control and situational circumstances as influences on behavioural intention rather than deterministic obstacles (Kamal et al., 2020).

The relatively low explanatory power of the regression model ($R\ 2 = 0.208$) also indicates that, although other factors such as user interface design, checkout ease, promotional incentives, and responsiveness of customer service are also factors in conversion behaviour. However, the most influential factor in the tested model will be customer trust, which will support its strategic role in Jumia's digital commerce operations. In practice, these findings imply that investments made without trust-building plans can yield minimal returns when the focus is on enhancing infrastructure.

6.0. Conclusion

The discussion examines the implications of mobile payment options for consumers' shopping preferences on Jumia in Nigeria. All in all, the deployment of mobile payments is very customer-trust-oriented because it has emerged as the most predictive of conversion rates. The findings show that customers are more likely to make purchases if they assume that Jumia's mobile payment systems are ready, transparent, and reliable. This aspect of trust not only boosts confidence in transactions but also the experience and overall interaction.

Concurrently, the infrastructure problems, such as the absence of internet connection and access to phones, were the manifestation of the weak direct impact on purchasing behaviour, which means that digitally savvy users of Jumia may be more vulnerable to the platform's safety and convenience rather than technical restrictions (Olarinde et al., 2024). The high trust, driven by improved security communication and a more user-friendly payment experience, will also maximise conversion performance, cementing Jumia as a robust digital commerce platform in Nigeria's rising mobile economy.

7.0. Recommendation

Based on the analysis and evaluation of the effect of mobile payment methods on customer decisions on Jumia's shopping platform in Nigeria, the study makes the following recommendations to improve the customers` experience by adopting the following approaches:

- *Strengthen Customer Trust:* Implement advanced security features, enhance payment transparency, and ensure responsive customer support to build confidence in Jumia’s mobile payment systems.

- *Communicate Security Measures Clearly:* Regularly inform users about data protection protocols to increase trust and reduce perceived transaction risks.
- *Optimise User Experience:* Simplify the mobile payment journey, improve interface design, and ensure quick, reliable transactions to enhance ease of use and satisfaction.
- *Leverage User Feedback:* Use customer reviews and behavioural analytics to identify friction points and continuously refine the mobile payment experience.
- *Monitor Infrastructure Readiness:* Maintain oversight of internet connectivity and smartphone accessibility trends to anticipate potential barriers to mobile payment adoption.
- *Promote Digital Confidence:* Educate customers on the safety and convenience of mobile payments to increase adoption and loyalty.

Theoretical and Practical Contributions

Theoretical Contribution

This study extends the mobile payment literature by shifting analytical focus from technology adoption intention to actual conversion behaviour within an operational e-commerce platform. While prior studies have largely emphasised behavioural intention and usage willingness, they have paid limited attention to transaction completion outcomes in platform-specific contexts (Khan et al., 2021; Putrevu & Mertzanis, 2024). By empirically demonstrating customer trust as a mediating mechanism, rather than merely an independent predictor, the study advances established frameworks such as the Technology Acceptance Model (TAM) and the Theory of Planned Behaviour (TPB), which have been criticised for under-theorising trust dynamics in high-risk digital payment environments (Kamal et al., 2020; Budiharseno & Kim, 2023). Furthermore, the provision of platform-level empirical evidence from Nigeria addresses a notable gap in emerging market research, which has predominantly relied on generalised or cross-country analyses rather than firm-specific behavioural outcomes (Olarinde et al., 2024).

Practical Contribution

From a managerial perspective, the findings offer actionable insights for Jumia and comparable e-commerce platforms by highlighting the strategic importance of payment user experience (UX), security transparency, and trust signalling in driving conversion rates. Consistent with prior digital commerce research, the results indicate that infrastructure investment alone is insufficient to improve transaction completion unless accompanied by deliberate trust-building initiatives, such as clear security communication and reliable transaction processes (Ankrah et al., 2024; Albshaier et al., 2024). This underscores the need for integrated platform strategies that combine technological reliability with effective communication and user assurance mechanisms to sustain consumer confidence and competitive performance in emerging digital markets.

References

Alabi, S. (2022). *Authentication technology methods for e-commerce applications in Nigeria—A case for biometric digital security contactless palm vein authentication* (Doctoral dissertation, University of Sussex).

Albshaier, L., Almarri, S., & Hafizur Rahman, M. M. (2024). A review of blockchain's role in e-commerce transactions: Open challenges and future research directions. *Computers, 13*(1), Article 27. https://doi.org/10.3390/computers13010027

Ali, I. M. (2024). A guide for positivist research paradigm: From philosophy to methodology. *Ideology Journal, 9*(2), 1–12.

Alturki, U., & Aldraiweesh, A. (2022). Adoption of Google Meet by postgraduate students: The role of task–technology fit and the TAM model. *Sustainability, 14*(23), Article 15765. https://doi.org/10.3390/su142315765

Anagreh, S., Al-Momani, A. A., Maabreh, H. M. A., Sharairi, J. A., Alrfai, M. M., Haija, A. A. A., & Al-Hawary, S. I. S. (2024). Mobile payment and digital financial inclusion: A study in the Jordanian banking sector using the unified theory of acceptance and use of technology. In *Business analytical capabilities and artificial intelligence-enabled analytics: Applications and challenges in the digital era* (Vol. 1, pp. 107–124). Springer Nature Switzerland.

Anjum, S., & Chai, J. (2020). Drivers of cash-on-delivery method of payment in e-commerce shopping: Evidence from Pakistan. *SAGE Open, 10*(3), 2158244020917392. https://doi.org/10.1177/2158244020917392

Ankrah, S. T., He, Z., Asare-Kyire, L., & Ofori, K. S. (2024). Beyond cash: A user-centric approach to mobile payment growth, service failure tolerance and continuance intention. *Total Quality Management & Business Excellence, 35*(15–16), 1847–1878. https://doi.org/10.1080/14783363.2022.2066787

Badran, M. F. (2021). Digital platforms in Africa: A case study of Jumia Egypt's digital platform. *Telecommunications Policy, 45*(3), Article 102077. https://doi.org/10.1016/j.telpol.2020.102077

Brandwijk, N. (2020). *The interplay of risks on digital platform openness: A case study* (Doctoral dissertation, Delft University of Technology).

Budiharseno, R. S., & Kim, M. J. (2023). Perceived privacy risks and trust dynamics: Rethinking mobile payment adoption in Indonesia. *Global Business & Finance Review, 28*(6), 112–129. https://doi.org/10.17549/gbfr.2023.28.6.112

Burgess, B., Yaoyuneyong, G., Pollitte, W. A., & Sullivan, P. (2023). Adopting retail technology in crises: Integrating TAM and prospect theory perspectives. *International Journal of Retail & Distribution Management, 51*(7), 939–954. https://doi.org/10.1108/IJRDM-08-2022-0369

Dhar, B. K., Omar, M. A., Yusuf, A. H., Oyelakin, I. O., & Mohamed, K. (2025). Ethical imperatives in digital finance: Advancing e-payment inclusion in a fragile economy. *Business and Society Review*.

Ezuwore-Obodoekwe, C. N., Eyisi, A. S., Emengini, S. E., & Chukwubuzo, A. F. (2014). A critical analysis of cashless banking policy in Nigeria. *IOSR Journal of Business and Management, 16*(5), 30–42.

Famosaya, O. B. (2024). *An assessment of the infrastructural readiness for transitioning to a cashless society in Nigeria* (Doctoral dissertation, National College of Ireland).

Ghali, Z. (2024). Impact of socio-economic status on customer e-loyalty under the moderating role of perceived self-efficacy. *Journal of Decision Systems, 33*(1), 53–78. https://doi.org/10.1080/12460125.2023.2241894

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018). On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. *Journal of Management Information Systems, 35*(1), 220–265. https://doi.org/10.1080/07421222.2018.1440766

Hagger, M. S., Cheung, M. W. L., Ajzen, I., & Hamilton, K. (2022). Perceived behavioural control moderating effects in the theory of planned behaviour: A meta-analysis. *Health Psychology, 41*(2), 155–167. https://doi.org/10.1037/hea0001120

Ifechukwu, A. (2022). *Regulating fintech in developing economies: Examining the risks, policies and Nigeria's path to financial prosperity*.

Islam, S. (2024). Impact of online payment systems on customer trust and loyalty in e-commerce: Analysing security and convenience. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5064838

Kamal, S. A., Shafiq, M., & Kakria, P. (2020). Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). *Technology in Society, 60*, Article 101212. https://doi.org/10.1016/j.techsoc.2019.101212

Khan, N. A., Khan, A. N., Bahadur, W., & Ali, M. (2021). Mobile payment adoption: A multi-theory model, multi-method approach and multi-country study. *International Journal of Mobile Communications, 19*(4), 467–491.

Khrais, L. T. (2020). Role of artificial intelligence in shaping consumer demand in e-commerce. *Future Internet, 12*(12), Article 226. https://doi.org/10.3390/fi12120226

Lindheim, M. B. T., & Grimsrud, O. M. (2017). *Merchant adoption of mobile financial services in Myanmar* (Master's thesis, Norwegian University of Science and Technology).

Lottu, O. A., Abdul, A. A., Daraojimba, D. O., Alabi, A. M., John-Ladega, A. A., & Daraojimba, C. (2023). Digital transformation in banking: A review of Nigeria's journey to economic prosperity. *International Journal of Advanced Economics, 5*(8), 215–238.

Mohapatra, A. G., Mohanty, A., Mohanty, S. K., Mahalik, N. P., & Nayak, S. (2025). Personalisation and customer experience in the era of data-driven marketing. In *Artificial intelligence-enabled businesses* (pp. 467–511). Springer.

Nucciarelli, A., & Sadowski, B. (2018). Managing uncertainty in the digital economy: Strategic and policy lessons from European broadband development. *Telecommunications Policy, 42*(10), 889–901.

Olarinde, E. S., Idem, U. J., & Obieze, I. D. D. (2024). Analysis of electronic commerce for promoting sustainable development in Nigeria. In *Proceedings of the 2024 International Conference on Decision Aid Sciences and Applications* (pp. 1–6). IEEE.

Olarinde, E. S., Idem, U. J., & Obieze, I. D. D. (2024). Analysis of electronic commerce for the promotion of sustainable development in Nigeria: Addressing challenges and envisaging future prospects. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1–6). IEEE.

Opebiyi, F. M. (2022). *Regulating user interactions within the financial technology market: Cryptocurrencies in Nigeria* (Doctoral dissertation, University of Manchester).

Putrevu, J., & Mertzanis, C. (2024). The adoption of digital payments in emerging economies: Challenges and policy responses. *Digital Policy, Regulation and Governance, 26*(5), 476–500. https://doi.org/10.1108/DPRG-06-2023-0054

Rawat, P. (2024). Consumer perception and adoption of digital payment methods: A study on trust and security concerns. *Educational Administration: Theory and Practice, 30*(4), 6022–6029.

Statista. (2024). *E-commerce spending by category in Nigeria.* https://www.statista.com/statistics/1139840/e-commerce-spending-in-nigeria-by-category/

Statista. (2025). *E-commerce market forecast: Nigeria.* https://www.statista.com/outlook/emo/ecommerce/nigeria

Wang, T., Liu, T., & Zhu, H. (2024). Cybersecurity challenges in mobile payment systems: A case study of Alipay in Chinese cities. *Innovation in Science and Technology, 3*(1), 51–58.

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 86-103
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1818

Comparative Analysis of Traditional Machine Learning, Deep Learning, and Hybrid Ensemble Models for Anomaly Detection and Web Application Firewall Optimisation

Shiva Nezamzadeh & Dilek Celik
Northumbria University, London

Abstract
Anomaly detection is an important component of cybersecurity, particularly in safeguarding web application firewalls (WAFs) from malicious traffic. In this study, we perform a comparative analysis of three Machine Learning (ML) approaches: Random Forest (RF), Convolutional Neural Network (CNN), and a stacking ensemble combining RF and CNN with Logistic Regression (LR) as the meta-learner to explore the most effective approach for anomaly detection. To ensure a fair comparison, we trained all models under consistent preprocessing pipelines, including data class balancing using the SMOTE technique to address the common imbalance in attack data. The results of this study showed that the stacking ensemble outperformed the other models, achieving the highest accuracy (99.97%). The CNN model followed closely with comparable accuracy (99.94%), while also offering significant advantages in terms of computational efficiency and interpretability, particularly when supplemented with SHAP analysis. In contrast, the RF model achieved moderate accuracy (80.41%) but demonstrated strengths in interpretability and efficiency. These findings highlight that, with effective preprocessing, a standalone CNN can provide a practical and resource-efficient alternative to more complex ensemble models. The findings of this study highlight the importance of preprocessing in optimising model performance and propose CNN as a suitable solution for real-time cybersecurity applications. Future research should explore these models across diverse datasets, further investigate hybrid deep learning (DL) frameworks, and integrate advanced interpretability methods to enhance model transparency and trust in ML-based security systems.

Keywords: Machine Learning, Deep Learning, Hybrid Ensemble Model, Anomaly Detection, Web Application Firewall Optimisation
Wordcount: 232

1.0 Introduction

The role of anomaly detection using ML models in network traffic cannot be overstated, particularly in light of increasingly sophisticated cyberattacks that are testing conventional models of security. Indeed, it can be seen that there is an urgent requirement for intelligent security models that can adapt to changing security challenges. WAFs are also of critical utility in network security, as they continually filter HTTP traffic in search of malicious traffic that can compromise web applications. However, it must also be recognized that there are limitations in terms of security that are tied to the models of detection that are adopted. Indeed, there have been several suggestions that ML can be of significant utility in terms of making WAF more responsive, particularly in terms of unearthing sophisticated cyberattacks that are using zero-day exploits and hidden attack models that are not easily detectable using conventional models of security, as has been underscored in previous studies (Nassif et al., 2021; Nti et al., 2022; Alghanmi, Alotaibi, & Buhari, 2022); however, there are also certain limitations in terms of models that can yield more practical results, including ML, DL, and ensemble models.

This study examines the effectiveness of three ML model categories of traditional ML algorithms, DL models, and hybrid ensemble approaches to enhance anomaly detection within WAFs. Accuracy, precision, recall, F1-score, and other performance metrics including computational efficiency, feature significance, and interpretability are used to assess the models. By applying consistent preprocessing and evaluation criteria, the study seeks to reveal the optimal model for enhancing WAF performance. This work contributes to the field of study by providing a systematic, side-by-side comparison of various traditional ML, DL, and hybrid ensemble models under a unified preprocessing, feature-selection, and evaluation pipeline, facilitating fair and reproducible performance assessment.

The main aim of this research is to increase the security of web applications through ML. Specifically, it compares various model categories in their ability to distinguish between legitimate and malicious traffic, providing insights into their practicality for real-world implementation.

1.2. Research objectives:

- To distinguish the strengths and limitations of traditional ML, DL, and hybrid ensemble models for anomaly detection in WAF.
- To identify the most effective model to enhance WAF security.
- To assess how different anomaly detection techniques impact model performance and efficiency.

1.3. Research Questions:

- How do traditional ML, DL, and hybrid ensemble models differ in their capacity to detect anomalies in web application traffic?
- Which model offers the best balance of accuracy, computational efficiency, and interpretability for anomaly detection in WAF?

The remainder of this paper is organised as follows: Section 2 provides details about the literature review, outlining existing approaches, key challenges, and research gaps. Section 3 presents the methodology, including the dataset, preprocessing steps, model selection, and evaluation metrics. Section 4 provides the experimental setup and model performance results. Section 5 delivers a comparative analysis, discussing each model's strengths, limitations, and practical implications. Finally, Section 6 concludes this research and offers recommendations for future research.

2.0. Literature Review

2.1. Anomaly Detection

Anomaly detection has evolved from statistical techniques to adaptive ML methods that can identify previously unseen attacks (Kalariya, Jethva and Alginahi, 2024). By implementing these techniques, Betarte, Pardo and Martinez (2018) proposed a model that uses HTTP header and payload attributes, achieving 98.4% accuracy on the OWASP dataset, highlighting the importance of resilient detection protocols.

2.2 Traditional Machine Learning Models and Their Ensembles in WAFs

Traditional ML models—such as RF, LR, Support Vector Machines (SVM), and Gradient Boosting—have been widely used in WAFs detect threats due to their simplicity and interpretability. SVM performs very well with high-dimensional data but may become inefficient with larger datasets. RF is a powerful method for network traffic classification, and it is robust against overfitting and high accuracy; however, it requires considerable computational resources, which may impact real-time applicability (Alserhani and Aljared, 2023; Acito, 2023). Ensemble models such as RF and Gradient Boosting can further improve accuracy by reducing false positives (FPs), which makes them suitable for WAFs. Gradient Boosting improves performance with sequential optimisation, but it is prone to overfitting and requires longer training times (Athief, Kishore and Paranthaman, 2024; Acito, 2023). Despite such limitations, ensemble approaches often outperform individual models, which are more vulnerable to high FP rates (Alserhani and Aljared, 2023).

Bagging, Stacking, and AdaBoost are common ensemble methods which improve classification performance by emphasising misclassified instances, a property particularly suited to binary anomaly detection tasks. However, AdaBoost can be sensitive to noisy data and outliers, which may negatively affect performance in real-world network traffic (Jeffrey et al., 2024; Odeh and Taleb, 2024). Advanced ensemble approaches combining multiple models have demonstrated strong performance. For instance, Tama et al. (2020) proposed a stacked ensemble integrating RF, Gradient Boosting Machine, and XGBoost, which achieved improved accuracy and reduced false positives across two datasets using grid search and k-fold cross-validation. Although computationally intensive, such approaches can deliver substantial performance gains.

Overall, traditional ML models such as RF, SVM, and LR continue to play a key role in web security due to their interpretability and efficiency. While these models face challenges with high-dimensional or large-scale data and potential overfitting, ensemble methods help mitigate many of these limitations. Nonetheless, ensuring robustness and effective integration across diverse models remains a challenge in ensemble design. These characteristics of traditional ML models and their ensemble strategies are summarised in Figure 1.

Figure 1. Traditional ML (Conceptual diagram and annotations created by the authors; embedded image reproduced from CrowdforThink, 2019*)*

2.3 DL Approaches and Their Ensembles in WAFs

DL models have become more important in improving WAFs. Tian et al. (2020) suggested a distributed DL framework for detecting online attacks on resource-limited edge devices. They showed that DL can support scalable and real-time threat detection while maintaining high accuracy. Among widely implemented DL models, Long Short-Term Memory (LSTM) networks have demonstrated strong performance in capturing long-term sequential patterns. However, despite achieving high accuracy in DDoS detection (97.57%), LSTM models are prone to overfitting and require substantial computational power, which can limit their practicality in real-time WAF implementations (Dawadi, Adhikari and Srivastava, 2023; Ali et al., 2022).

CNN is another DL model that shows strong relevance to cybersecurity. Although CNNs were originally developed for image processing, they have been effectively used for malware and network traffic analysis by automatically learning discriminative characteristics. Nevertheless, their high computing requirements can limit suitability for real-time WAF deployment, especially when detecting infrequent attack patterns (Ali et al., 2022; Kimanzi et al., 2024).

Although DL ensembles have shown performance gains in WAF-based intrusion detection, their effectiveness closely depends on careful model selection and integration. For instance, PANACEA—an ensemble that combined CNNs, RNNs, and Autoencoders—achieved high detection accuracy but faced difficulties in identifying the dominant contributing model. Similarly, ResNet-18 has reported standalone classification accuracy of around 77%, which highlights the significance of selecting robust base models when designing ensemble architectures (AL-Essa et al., 2024; Tan, 2023; Li et al., 2023; Alanazi et al., 2022).

To manage the complexity of DL ensembles, various aggregation strategies have been proposed; model performance and computing limitations are usually used to guide selection (Waheed et al., 2023). Average voting, majority voting, and optimal weighting are common approaches. Average voting is frequently stated to deliver more stable performance by balancing model predictions, while majority voting offers simplicity at the cost of decreased sensitivity. By giving each model a certain level of priority, optimal weighting can increase accuracy even more, although it adds more computing overhead (Alanazi et al., 2022).

In summary, DL models such as LSTM and CNN have enhanced WAF capabilities by achieving high detection accuracy. However, their practical deployment in real-time environments can be constrained by overfitting risks and computational demands. While DL ensembles can further improve performance, they introduce additional complexity in model selection and system integration. These strengths and limitations of DL approaches are summarised in Figure 2.

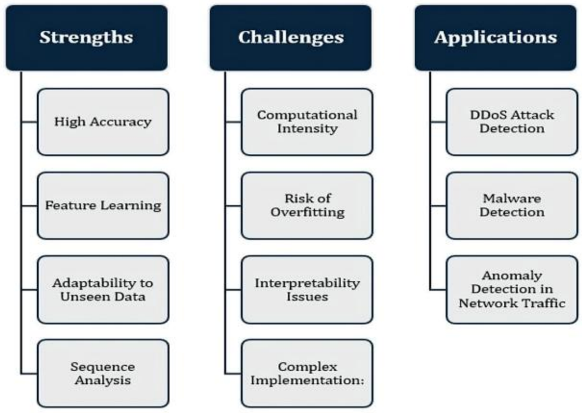

Figure 2. DL approaches (Conceptual diagram created by the authors)

2.4 Ensemble Techniques combining Traditional and DL models

Hybrid ensemble methodologies, which merge conventional ML and DL models, leverage the synergistic benefits inherent in both approaches, demonstrating considerable promise for anomaly detection within WAFs. Although DL models possess robust feature-learning capabilities, they frequently exhibit high computational demands. Traditional ML models, including RF, Gradient Boosting, and SVM, can augment DL models by providing enhanced efficiency and interpretability. Through feature-level or decision-level fusion, hybrid ensembles can attain elevated classification accuracy while addressing the constraints of individual models (Tan, 2023; Ovi, Rahman and Hossain, 2024).

A pertinent illustration is the hybrid ensemble introduced by Abdelmounaim and Madani (2024), which integrated traditional ML models (XGBoost and RF) with DL models (CNN and RNN) employing a stacking architecture. Their methodology exhibited superior classification accuracy compared to standalone models, thereby indicating improved robustness and adaptability within intrusion detection applications.
Notwithstanding these advantages, hybrid ensembles present increased complexity. The heightened model heterogeneity introduces difficulties in training, tuning, and deployment, especially within real-time or resource-limited settings (Xu et al., 2024). Moreover, ensemble strategies like stacking and blending could potentially amplify the vulnerabilities of individual models if the integration process is not meticulously planned (Li et al., 2023).

Overall, hybrid ML–DL ensemble models provide significant performance enhancements for cybersecurity applications; however, challenges concerning interpretability, scalability, and deployment intricacy could restrict their practical utility in specific WAF implementations.

2.5 Research Gap

This review of existing research critically examines important approaches to anomaly detection using ML, including traditional ML algorithms, DL architectures, and combined ensemble methods. While previous research has studied the strengths and weaknesses of these methods separately, there is a lack of research that systematically compares all three within a single experimental setup for WAF optimisation (Xu et al., 2024; Alanazi et al., 2022; Ali et al., 2022). This research gap makes it challenging to fully understand the trade-offs between prediction accuracy, computational efficiency, and practical use. To address this, this study gives a comparative evaluation of models from each category, using consistent preprocessing and evaluation methods. By clarifying the relative advantages of each method, this research aims to help choose scalable, reliable, and understandable ML solutions to improve WAF-based cybersecurity.

3.0. Research Methodology and Planning

3.1. Research Design

This study, which uses a quantitative and comparative approach, evaluates different ML models for finding anomalies, following the CRISP-DM framework. The research is based on positivism, assuming that how well models perform provides an objective view of reality. It also takes a realist approach, considering anomaly detection as something that can be observed and understood through data analysis. Additionally, the study is empirically driven, aiming to create knowledge through systematic experimentation and quantitative analysis. Methodologically, the research follows the CRISP-DM process, which includes five main stages: (1) Data Understanding, (2) Data Preparation, (3) Modelling, (4) Evaluation, and (5) Interpretation (Gill et al., 2023; Brodie, 2024). (Figure 3)

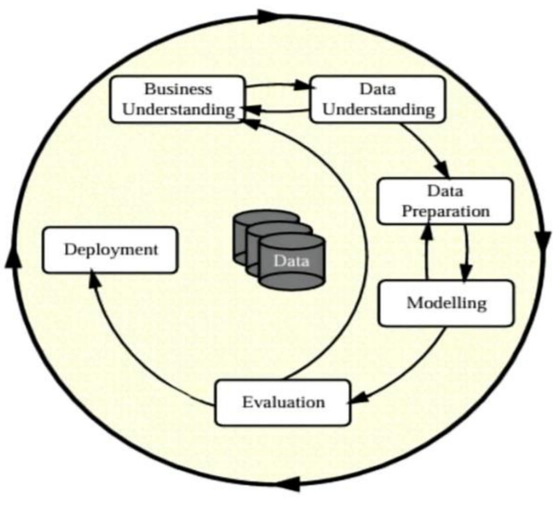

Figure 3. CRISP-DM process (Wirth and Hipp, n.d.)

3.2. Data Collection Methods and Data Understanding

This study employs the UNSW_NB15 dataset, a publicly accessible resource acquired from Kaggle, and extensively utilised in academic investigations pertaining to intrusion detection. The dataset offers distinct training and testing files, each encompassing in excess of 50,000 network traffic instances, thereby facilitating robust model training and evaluation for the purpose of anomaly detection. All records within the dataset are completely anonymised, ensuring the absence of any personally identifiable information. Given the absence of human subjects and the non-utilisation of sensitive personal data, prior ethical approval was deemed unnecessary for the present study.

3.3. Data Processing and Preparation

3.3.1 Dataset Overview

The dataset includes 45 features, including 41 numerical and 4 categorical attributes. The binary target variable (Label) denotes normal traffic (0) and attacks (1). No missing or duplicate values were found.

3.3.2 Data Engineering

One-hot encoding was used to represent categorical features, while numerical features were standardised using Z-score normalisation.

3.3.3 Outlier Handling and Correlation Analysis

Initial exploratory analysis included visual inspections (histograms, boxplots, scatter plots) and correlation assessment to understand feature behaviour. Since the dataset contains non-Gaussian distributions and a mixture of linear and non-linear relationships, model-based methods were prioritised for robust outlier handling and variable relevance assessment.

3.3.4 Model-based feature selection

To improve model performance and address dataset complexity, RF and XGBoost were employed for feature selection and outlier detection. Feature importance scores from these models were used to rank attributes, with attack_cat_Normal identified as having the strongest influence on the target variable (Label). Instances misclassified by either model were flagged as potential outliers. Crucially, no extreme values were removed, capped, or otherwise altered, as they were considered intrinsic to the dataset and representative of authentic network behaviour. Preserving these extreme values allows the models to learn from the complete spectrum of normal and malicious traffic, which is essential for realistic intrusion detection. Combined with model-driven feature selection, this approach reduced dataset dimensionality and optimised it for subsequent modelling. To address class imbalance, SMOTE and undersampling techniques were used. This study combined RF and XGBoost with both SMOTE and undersampling to create a balanced dataset optimised for anomaly detection (Vibhute et al., 2024; Pansari et al., 2024; Putra, 2024).

3.4. Data Analysis Methods

The main objective of this study is to identify the most suitable model from each category (traditional ML, DL, and hybrid ensembles). Only ensemble-based models were selected from the ML and DL categories to avoid bias from comparing single models to ensembles. Collectively, these models were chosen to represent complementary strengths across interpretability (RF), deep feature learning (CNN), and heterogeneous model integration (stacking), thereby ensuring a systematic comparison across traditional ML, DL, and hybrid ensemble paradigms. RF, an ensemble CNN, and a stacking approach were chosen as representative models based on their characteristics, computational requirements, and suitability for the dataset. RF was used as the traditional machine learning ensemble baseline due to its robustness to noisy and high-dimensional data, strong resistance to overfitting, interpretability, and comparatively low computational complexity (Panasov and Nechitaylo, 2021), making it a well-established benchmark in intrusion detection research.

An ensemble CNN was chosen to represent DL-based methods because of its ability to automatically learn hierarchical and non-linear feature representations, while ensemble aggregation enhances generalisation performance. CNN ensembles have consistently achieved high detection accuracy and demonstrated strong scalability in network intrusion detection, although they are associated with higher computational intensity (Yang, Lv and Chen, 2022). Stacking was employed as a hybrid ensemble learning representative to explicitly evaluate whether combining heterogeneous models with fundamentally different architectures can enhance detection performance. Unlike voting or boosting techniques, stacking enables a meta-learner to learn optimal combinations of base model predictions. In this study, logistic regression (LR) was adopted as a lightweight meta-learner to minimise additional complexity while integrating RF and CNN outputs (Zhang and Wang, 2023).

Other candidate models, such as support vector machines, gradient boosting techniques, and recurrent neural networks, were excluded because they either conceptually overlapped with the selected approaches or relied on temporal assumptions outside the scope of this study. This selection strategy enables a focused yet representative comparison across traditional ML, DL, and hybrid ensemble paradigms for intrusion detection (Azam et al., 2023; Javaid et al., 2016).

3.5. Validation and Testing

Since the dataset provided distinct training and testing files, data splitting was inherently addressed, and preprocessing was performed separately on each subset. Stratified K-Fold Cross-Validation was used during model training to preserve class balance and reduce overfitting (Chen et al., 2023). Hyperparameter tuning was conducted using GridSearchCV, which also incorporated stratified folds. Although computationally intensive, Grid Search was selected for its reliability in cybersecurity contexts, where accuracy is critical. Given the computational demands of the CNN model, a reduced hyperparameter grid was applied. This ensured adequate optimisation without imposing excessive resource demands (Masum et al., 2021; Franceschi et al., 2024).

Model performance was evaluated using a comprehensive set of metrics to ensure balanced and reliable comparisons. Accuracy, Precision, Recall, and F1-Score were used to assess predictive performance. These metrics capture the trade-off between FPs and false negatives (FNs), which is crucial in anomaly detection (Yang et al., 2023; Li et al., 2023). ROC-AUC was applied to evaluate the models' ability to discriminate across thresholds. Balanced Accuracy was included to address class imbalance and ensure fair evaluation across both classes (Owusu-Adjei et al., 2023). To strengthen reliability, agreement metrics such as Matthews Correlation Coefficient (MCC) and Cohen's Kappa were incorporated (Chicco and Jurman, 2020). Additional error-based metrics including Specificity, Fall-out (FPR), Negative Predictive Value (NPV), False Discovery Rate (FDR), and False Omission Rate (FOR) were used to capture the impact of misclassification (Larner, 2024).

Finally, confusion matrices were generated to provide a clear breakdown of classification outcomes. To assess computational efficiency and interpretability, inference speed, memory usage, and SHAP values were included (Wang and Wang, 2022). RF was further assessed through model complexity, permutation importance, and feature importance (Ahsan et al., 2021), and model size was specifically measured for CNN (Saleem et al., 2022). Results were visualised to provide a clear comparison of overall model efficiency.

4.0. Results and Analysis

4.1. Explanatory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to identify key patterns in the dataset and guide subsequent modelling decisions. The distribution of the Label variable indicated a clear class imbalance, with malicious traffic dominating in both the training and test datasets (Figure 4).

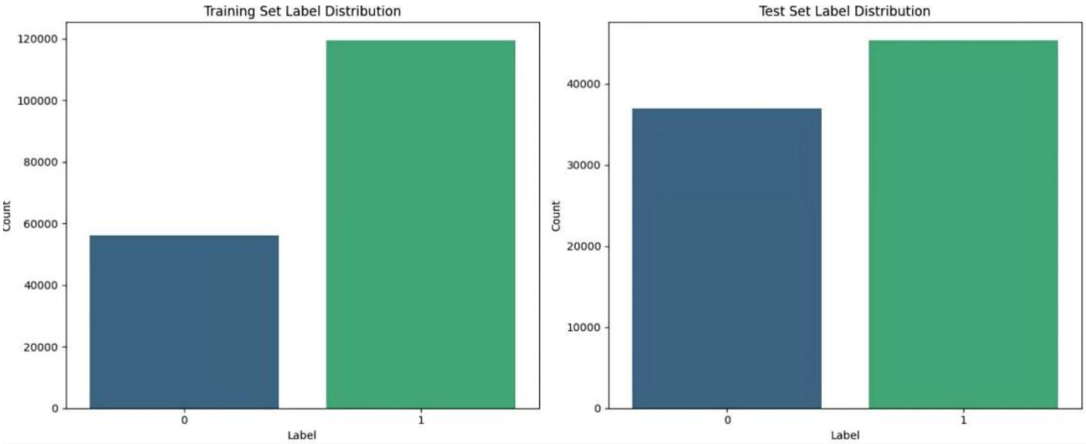

Figure 4. Label distribution of the training and test datasets

Correlation analysis showed that sttl (source time-to-live) and ct_dst_sport_ltm (count of connections to the destination over time) had the strongest positive associations with the Label variable. In contrast, swin (source window size) and dload (data transmission rate from destination to source) displayed the strongest negative correlations (Figure 5). Similar correlation trends were observed in the test dataset, indicating consistent feature behaviour across data splits.

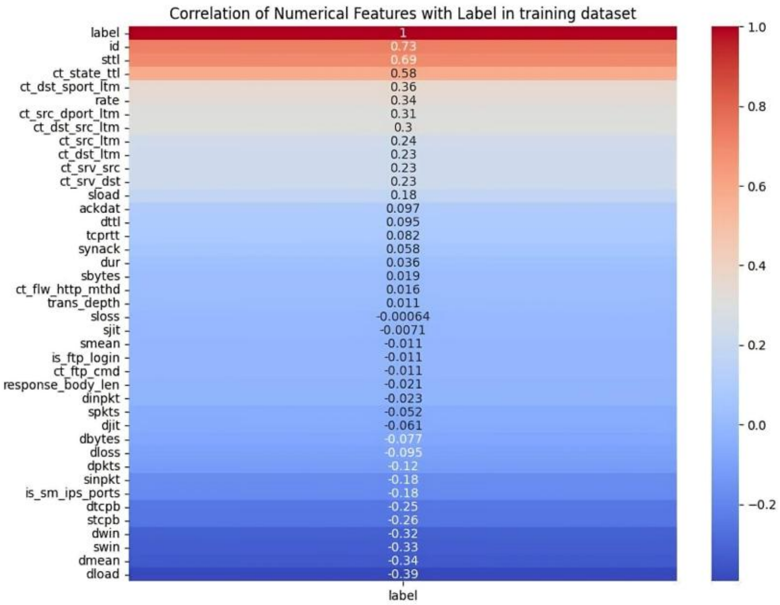

Figure 5. Correlation of 'Label' with other features in the training dataset.

Additional analyses, including service type distributions and attack category comparisons, were performed to further characterise the dataset. Overall, the EDA identified meaningful patterns and confirmed the dataset's complexity, underscoring the need for robust preprocessing prior to model training.

4.2 Outlier handling, Correlation Analysis, Model-based feature selection

4.2.1 Outlier Detection, Distribution, and Correlation Analysis

Exploratory analysis using histograms, boxplots, and scatter plots indicated that many numerical features deviated from normality, with substantial skewness and kurtosis. This observation motivated the use of non-parametric methods in subsequent analysis. Figure 6 presents a representative example using the spkts feature (the number of packets transmitted from the source during a network flow), which demonstrates a highly skewed distribution. Given these characteristics, Spearman's rank correlation was applied to examine relationships between numerical features and the target variable (Label), revealing varying degrees of association. For categorical variables, ANOVA tests were used to identify both significant and non-significant relationships. These findings informed the feature selection process and subsequent model design.

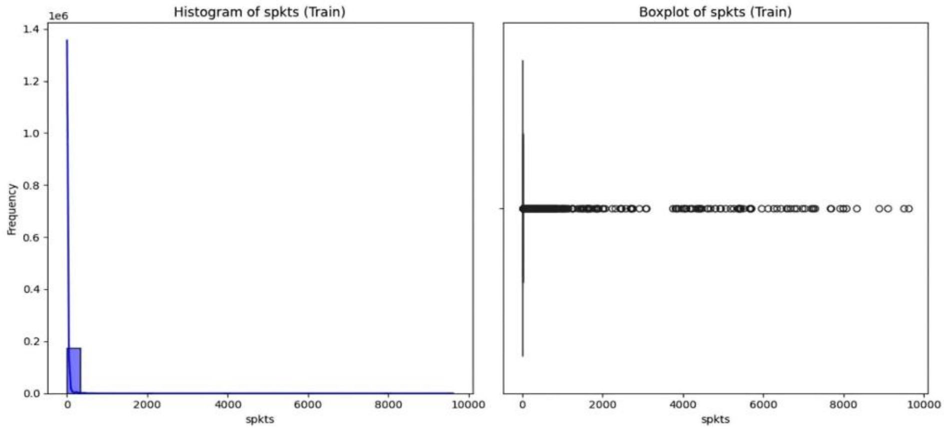

Figure 6. Histogram and Boxplot of 'spkts' in the training dataset

4.2.2 Model-based feature selection
RF and XGBoost were utilised for feature selection and outlier analysis to improve model performance and manage dataset complexity. Feature importance rankings from these models were used to prioritise the most influential attributes, with attack_cat_Normal identified as the most dominant predictor of the target variable (Label). Potential outliers were identified through the models' misclassification signals. Since neither model flagged extreme values for removal, all observations were retained, suggesting that the extreme values are intrinsic to the dataset. This approach reduced dataset dimensionality and prepared the data for subsequent modelling. Furthermore, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training dataset to address class imbalance.

4.3 Models Application
4.3.1 Random Forest
The RF model was optimised using GridSearchCV with Stratified K-Fold Cross-Validation, as outlined in Section 3. It achieved an overall accuracy of 80.4% and a balanced accuracy of 78.4%, indicating consistent performance despite class imbalance. In cybersecurity contexts, such results constitute a reasonable baseline, although recall and discrimination capability are more critical than accuracy alone. The RF model achieved a recall of 80.4% and a ROC–AUC score of 0.94, which demonstrated strong power to distinguish between normal and anomalous traffic (Table 1). Confusion matrix analysis (Table 2) confirms effective identification of attack instances with a relatively low number of false negatives, although some benign traffic was misclassified.

Table 1. RF model evaluation metrics

Metrics	Value
Accuracy	0.8041
Balanced Accuracy	0.7842
Precision	0.8422
Recall	0.8041
F1- Score	0.7939
MCC	0.6336
Cohen's Kappa	0.5895
Specificity	0.5873
ROC AUC Score	0.9420
Sensitivity (TPR)	0.8041
FPR	0.4127
NPV	0.9620
FDR	0.2556
FOR	0.0380

Table 2. RF model confusion matrix

Metrics	Value
TP	44,473
TN	21,731
FP	15,269
FN	859

From a computational perspective, the RF model achieved high inference speed and low memory consumption (Table 3), owing to its shallow tree structure. This highlights its suitability for real-time or resource-constrained deployment scenarios.

Table 3. RF model computational metrics

Metric	Value
Inference Speed	6,305,931.79 samples per second
Memory Usage	491.4296875
Number of Estimators	5
Maximum Depth	3
Maximum Features	sqrt

Model interpretability was examined by using built-in feature importance, permutation importance, and SHAP analysis. Across all approaches, sttl and attack_cat_Normal emerged as the most influential features, indicating transparent model behaviour. The SHAP interaction analysis further illustrated the combined influence of these features on prediction outcomes (Tables 4–5; Figure 7).

Table 4. RF model feature importance based on the built-in feature importance

Feature	Importance
Sttl	0.202839
Ct_state_ttl	0.184023
Sload	0.155977
Dload	0.127539
rate	0.106322

Table 5. RF model feature importance based on permutation

Feature	Permutation Importance
Attack_cat_Normal	0.112580
State_INT	0.032521
Smean	0.017683
Tcprtt	0.015555
demean	0.012592

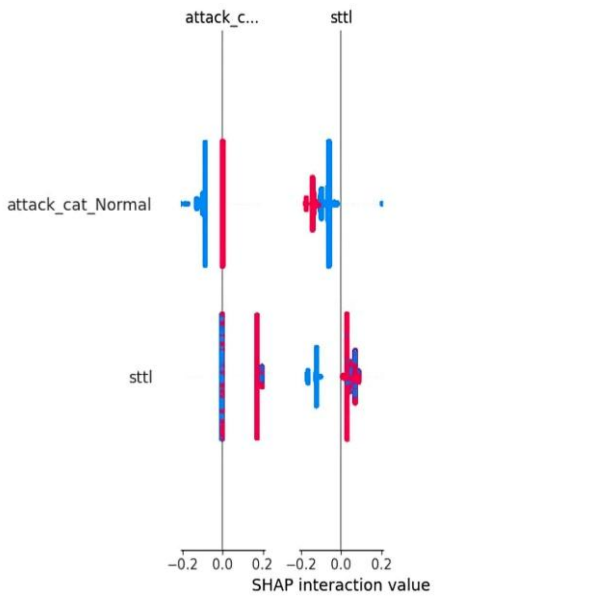

Figure 7. RF model SHAP interaction plot

4.3.2 CNN

To ensure reproducibility, random seeds were fixed for both NumPy and TensorFlow, and deterministic data splitting was applied throughout the experimental pipeline. The CNN model operated on a reshaped onedimensional feature vector using a single Conv1D layer with 32 filters and a kernel size of 3, followed by batch normalisation and max pooling with a pool size of 2. A global average pooling layer was then applied to summarise feature maps and reduce model complexity. The extracted features were passed to a fully connected dense layer with eight neurons and ReLU activation, followed by a dropout layer with a rate of 0.5 for regularisation. The final classification layer used a softmax activation function.

Hyperparameter tuning was performed using stratified three-fold cross-validation to evaluate combinations of learning rate, number of filters, and dropout rate. Learning rate 0.0005, 32 filters, and dropout rate 0.5, which were selected based on mean cross-validation accuracy, were the optimal configuration. The model was subsequently trained using an 80/20 training–validation split for 10 epochs. Analysis of the learning curves showed close alignment between training and validation accuracy and loss, indicating stable convergence and no evidence of overfitting. (Figure 8)

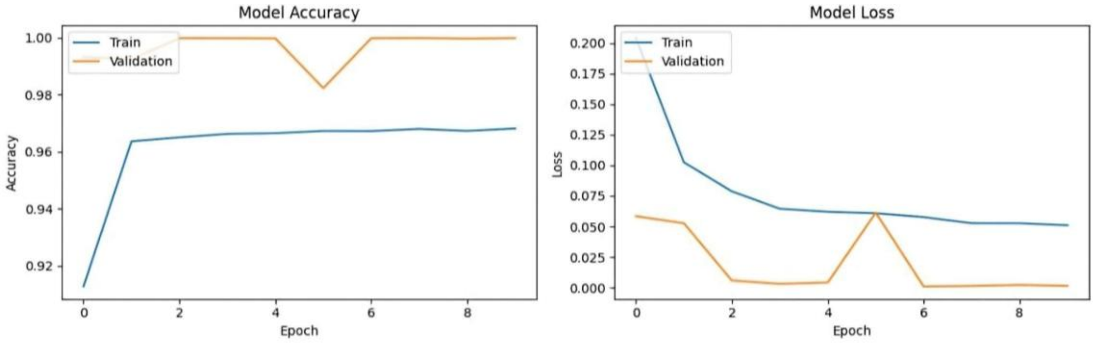

Figure 8. Accuracy and Loss of the train and validation set over 10 epochs

The CNN model demonstrated consistently high performance across all evaluation metrics, with accuracy, precision, recall, and F1-score all exceeding 99% (Table 6). The balanced accuracy of 99.81% and ROC–AUC of 0.9999 indicate excellent discrimination between normal and attack traffic. Confusion matrix results (Table 7) reveal minimal false positive and false negative rates, which highlight the model's robustness in detecting both attack and benign traffic.

Table 6. CNN model evaluation metrics

Metrics	Value
Accuracy	0.9994
Balanced Accuracy	0.9981
Precision	0.9994
Recall	0.9994
F1- Score	0.9994
MCC	0.9963
Cohen's Kappa	0.9963
Specificity	0.9993
ROC AUC Score	0.9999
TPR	0.9994
FPR	0.0007
NPV	0.9993
FDR	0.0006
FOR	0.0007

Table 7. CNN model confusion matrix

Metrics	Value
TP	45,306
TN	36,973
FP	27
FN	26

The CNN model demonstrated inference performance suitable for large-scale data processing. While its memory usage was higher than that of traditional ML models, the relatively small model size reflects an efficient architecture given its strong predictive performance (Table 8).

Table 8. CNN model computational metrics

Metric	Value
Inference Speed	19311.66
Average Memory Usage	5297.43
Model Size (MB)	0.05

SHAP analysis was used to examine feature contributions. The results identified *sttl* and *attack_cat_Normal* as the most influential features shaping the CNN model's predictions (Table 9; Figure 9).

Table 9. CNN model feature importance based on SHAP analysis

Feature	Importance Value
Sttl	0.174911
Attack_cat_Normal	0.173987
Ct_state_ttl	0.160816
Dload	0.160719
Attack_cat_Fuzzers	0.082288

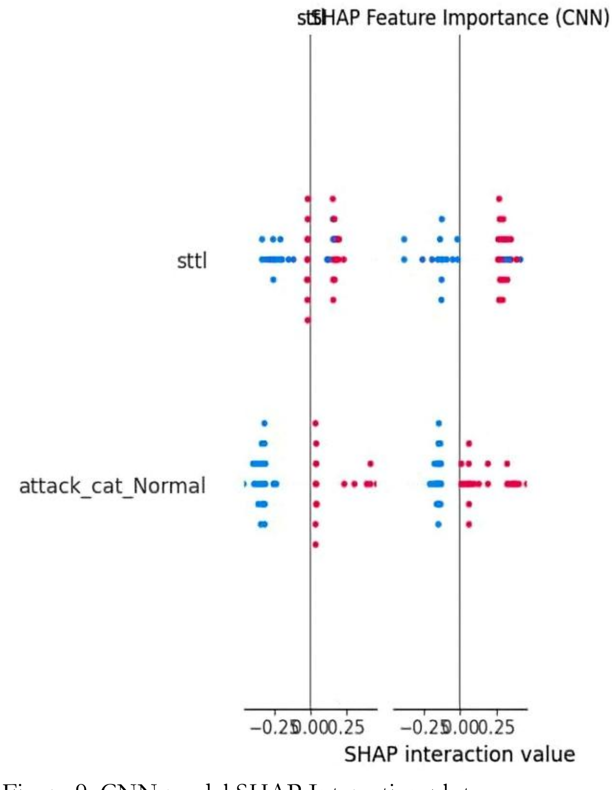

Figure 9. CNN model SHAP Interaction plot

4.3.3 Stacking hybrid ensemble

The stacking hybrid ensemble combined two trained base models: an RF, which generated class probabilities, and a CNN, which produced predictions. These outputs were combined into a new feature set used to train a lightweight LR meta-model. The ensemble delivered near-perfect classification performance, with accuracy, precision, recall, and F1-score all reaching approximately 99.97% (Table 10). Moreover, high MCC and Cohen's Kappa values indicate excellent agreement and robustness. As shown in the confusion matrix (Table 11), misclassification rates were minimal, with only 5 false positives and 21 false negatives, confirming the ensemble's detection capability.

Table 10. Hybrid Ensemble evaluation metrics

Metrics	Value
Accuracy	0.9997
Balanced Accuracy	0.9997
Precision	0.9997
Recall	0.9997
F1- Score	0.9997
MCC	0.9994
Cohen's Kappa	0.9994
Specificity	0.9999
ROC AUC Score	0.9997
TPR	0.9995
FPR	0.0001
NPV	0.9994
FDR	0.0001
FOR	0.0006

Table 11. Hybrid Ensemble confusion matrix

Metrics	Value
TP	45,311
TN	36,995
FP	5
FN	21

The computational characteristics of the stacking ensemble were analysed to complement the performance evaluation (Table 12). The reported inference speed, memory usage, and model size reflect the computational behaviour of the ensemble architecture itself and should be interpreted as indicative rather than exhaustive, as the overall computational intensity depends on the underlying base models.

Table 12. Hybrid Ensemble computational metrics

Metric	Value
Inference Speed	60657902.15
Average Memory Usage	5885.76
Model Size	19.05

Permutation importance analysis was used to examine feature contributions within the stacking ensemble (Figure 10). The results indicate that the CNN-derived features dominate the ensemble's decision-making. This confirms that the ensemble's predictions are primarily driven by the CNN component rather than an equal contribution from both base models.

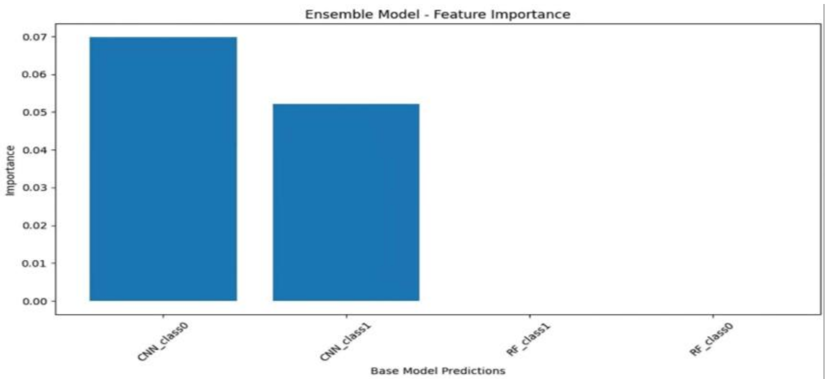

Figure 10. Hybrid Ensemble Feature Importance based on Permutation

5.0. Discussion and Implications

5.1 Model Performance

Figures 11–13 present a comparative analysis of the three evaluated models. Both the CNN model (accuracy: 99.94%) and the ensemble model (accuracy: 99.97%) illustrate similarly high levels of accuracy, substantially outperforming the RF model (accuracy: 80.41%). While the RF model achieved reasonably good performance, it was clearly less effective than the other two models.

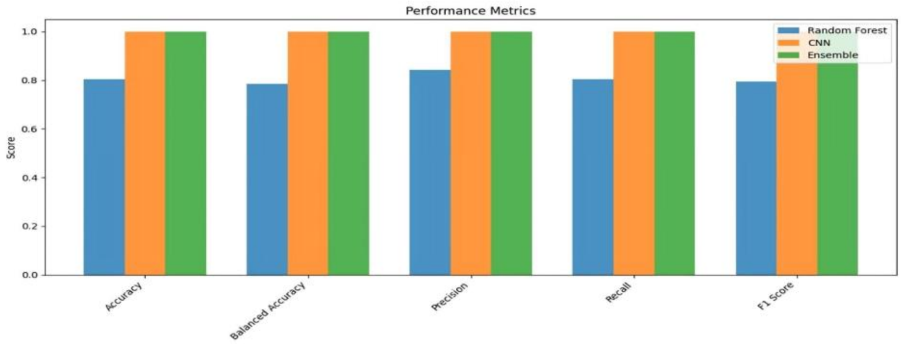

Figure 11. Comparison of models: Evaluation Metrics

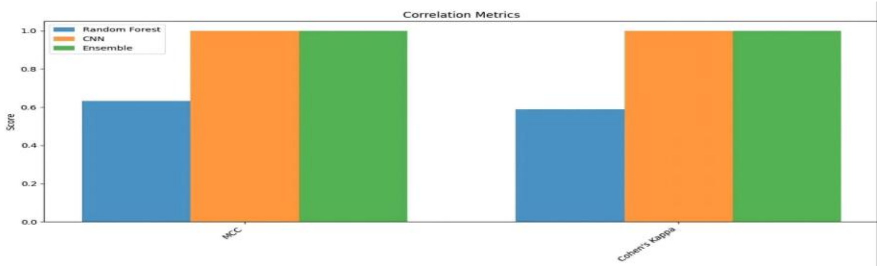

Figure 12. MCC and Cohen's Kappa comparison of models

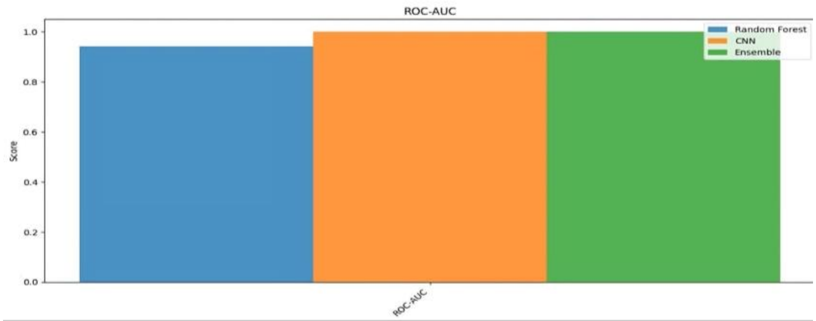

Figure 13. ROC-AUC comparison of models

Table 13 summarises the computational intensity metrics for each model. It is important to note that the reported inference speed of the ensemble model may not accurately reflect its true computational cost, as it is influenced by the combined performance of its constituent models. With this consideration, the ensemble model exhibits the highest computational intensity, likely due to its layered architecture. The CNN model shows moderate computational demand, while the RF model remains the most computationally efficient model.

Table 13 Comparison of computational metrics of models

Model	Inference Speed	Average Memory Usage	Model Size	Additional Details
RF	6,305,931.79	491.43	N/A (Tree-based)	5 estimators, max depth 3
CNN	19,311.66	5,297.43	0.05	N/A
Ensemble	60,657,902.15	5,885.76	19.05	These results should be interpreted alongside the computational characteristics of the RF and CNN base models

The overview of the feature importance and model interpretability is provided in Table 14. As shown, *'sttl'* and *'attack_cat_Normal'* emerged as the most influential features in both the RF and CNN models. The RF model is inherently interpretable, owing to its tree-based structure, allowing for clear insight into feature contributions. In contrast, the CNN model's deeper architecture introduces challenges in interpretability. The stacking ensemble model further increases the model complexity by integrating multiple learners. Nonetheless, analysis revealed that the CNN was the dominant model driving the ensemble's predictions.

Table 14. Comparison of feature importance and interpretability of models

Model	Key Contributing Features	Interpretability Assessment
RF	'sttl', 'attack_cat_Normal'	High, due to the intrinsic feature-based structure
CNN	'attack_cat_Normal', 'sttl'	Moderate; relies on complex internal representations
Ensemble	'CNN_class0', 'CNN_class1'	Moderate; interpretable via constituent models' contributions

A unified preprocessing and feature extraction was applied across all models, including RF, CNN, and a stacking ensemble with LR as the meta-learner. While both the CNN and the ensemble models demonstrated strong and comparable predictive performance, the RF model delivered comparatively lower results. Feature importance analysis confirmed the substantial contribution of the CNN model to the ensemble's performance. In terms of computational efficiency, the RF model required the lowest computational resources while offering the highest interpretability; by contrast, the CNN and ensemble models demanded greater computational costs, with the ensemble being the most resource-intensive. Despite the inherent complexity of the CNN model, its interpretability can be enhanced through post hoc explanation methods. Overall, the CNN model emerged as the most practical and effective model within the scope of this research.

5.2 Comparison with previous studies

This study employed a unified preprocessing and feature selection pipeline combining XGBoost, RF, and SMOTE across all models to ensure a fair comparison. Inconsistent preprocessing can inflate model-specific performance while weakening comparative validity, whereas systematic pipelines improve generalisation and interpretability (Farha and Ahmed, 2024; Ahsan et al., 2021). Consistent with these findings, Vibhute et al. (2024) demonstrated that RF-based feature selection significantly enhances CNN performance on the UNSW-NB15 dataset, achieving accuracy comparable to more complex ensemble models. Effective preprocessing allows simpler architectures to rival computationally intensive ensembles, a conclusion reinforced by the strong performance of the standalone CNN in this study. Hybrid ensemble models are often employed to improve detection accuracy; however, they frequently bring higher computational cost. For example, Farha and Ahmed (2024) reported resource limitations in a stacking ensemble incorporating an FNN. In contrast, the present study demonstrates that a simpler CNN-based design, when supported by effective preprocessing, can deliver competitive performance without excessive computational demands.

5.3 Contribution to the field

This study advances cybersecurity research by clarifying the relationship between preprocessing strategies, model architecture, and computational efficiency. By standardising preprocessing and feature selection, the research ensured methodological consistency and reduced bias in model comparisons. The integration of traditional ML and DL within the ensemble framework demonstrated that high predictive performance can be achieved without reliance on a complex meta-model. Importantly, the results show that, with well-structured preprocessing and feature extraction, a single CNN model can match the performance of an ensemble model, while requiring lower computational intensity. The application of SHAP and permutation importance enhanced the interpretability of the models. Overall, these findings contribute to a more balanced understanding of how predictive accuracy, interpretability, and computational efficiency can be jointly optimised in the design of practical intrusion detection systems.

5.4 Implications for Business Digital Innovation and Cybersecurity Strategy

Prior research highlights the strategic role of artificial intelligence in strengthening organisational decision-making and digital resilience by improving risk awareness, supporting data-driven management, and guiding cybersecurity investment planning. From a business perspective, AI-driven cybersecurity improves operational efficiency, reduces financial losses and service disruptions, and strengthens regulatory compliance through explainable security mechanisms. Together, these benefits underscore the role of AI-enabled security solutions in enhancing digital service reliability, fostering organisational trust and long-term competitive advantage (Daram and Senthilkumar, 2025; Sissodia et al., 2025).

Building on this foundation, the findings of this study suggest that selecting an anomaly detection model should be viewed as a strategic business decision rather than a purely technical choice. Models that combine high detection accuracy with manageable computational requirements can reduce infrastructure and operational costs, making them suitable for long-term deployment. Concurrently, stronger detection performance reduces the likelihood of successful cyberattacks, thereby limiting financial losses, service downtime, and reputational damage. This research explicitly balances predictive accuracy, interpretability, and computational efficiency, offering organisations a practical framework for selecting intrusion detection solutions that align with budgetary constraints and risk tolerance. Such a balanced approach supports cost-effective cybersecurity investment and maximises the long-term return on investment in AI-enabled WAF solutions.

5.5. Limitations

Despite the strong performance reported in this study, several limitations should be acknowledged. First, the experimental evaluation was conducted using a single intrusion detection dataset. Although UNSW-NB15 is widely used, model performance may vary across datasets with different traffic characteristics and attack distributions, potentially limiting generalisability.

Second, the results are influenced by the available computational resources. Variations in hardware can affect training efficiency, model complexity, and, to a limited extent, predictive performance; therefore, the findings should be interpreted within the context of the specific experimental setup used in this study.

Third, while representative models from traditional ML, DL, and hybrid ensemble categories were deliberately selected to enable systematic comparison, this choice may introduce methodological bias. Alternative architectures or more complex model configurations may achieve better performance, but their inclusion was constrained by practical computational considerations.

Finally, the exceptionally high accuracy achieved by the CNN and stacking ensemble should be interpreted with caution. Feature importance and SHAP analyses consistently identified *attack_cat_Normal* as a highly influential feature, which may contribute to optimistic performance estimates. This effect was explicitly analysed to ensure transparency. Accordingly, the primary contribution of this study lies in the controlled and fair comparison of traditional ML, DL, and hybrid ensemble models under a unified preprocessing framework, highlighting practical trade-offs between accuracy, interpretability, and computational cost.

6.0. Conclusions and Recommendations

6.1 Conclusion

This study shows that the CNN model achieved the highest overall performance for anomaly detection, with the stacking ensemble delivering comparable accuracy but at a significantly higher computational cost. The RF model indicated acceptable performance but remained less accurate than the DL approaches. Feature importance analyses revealed 'sttl' and 'attack_cat_Normal' as key predictors in both the CNN and RF models, with the ensemble's decisions primarily influenced by the CNN component. Consistent preprocessing using XGBoost and RF for feature extraction and SMOTE for class balancing was critical to achieving fair comparisons and strong model performance. These preprocessing steps enabled the CNN model to perform comparably to more complex models, confirming that effective data preparation can reduce the reliance on computationally intensive designs. Overall, these findings support the selection of intrusion detection solutions that balance detection effectiveness with computational efficiency, enabling sustainable and cost-conscious deployment in real-world environments.

6.2 Recommendation for future findings

Based on the findings of this study, several directions for future research can be identified. First, validating the proposed models on multiple and more diverse intrusion detection datasets would help assess their robustness and improve generalisability across different real-world network environments.

Second, future work could explore more advanced deep learning architectures. Integration of multiple deep learning techniques within a unified framework, such as dual CNN structures or CNN-RNN combinations, may enhance detection performance while maintaining acceptable computational efficiency.

Finally, improving model interpretability represents an important avenue for future study. Investigating advanced explainability methods, including saliency maps, Layer-wise Relevance Propagation (LRP), and attention-based visualisation techniques, could enhance model trustworthiness and support practical deployment (Ahmed and Jalal, 2024; Achtibat et al., 2024; Wang, Ouyang and Zeng, 2024).

References

Abdelmounaim, K. & Madani, M.A. (2024) 'A hybrid ensemble approach integrating machine learning and deep learning with sentence embeddings for webpage content classification', in *Modern Artificial Intelligence and Data Science 2024: Tools, Techniques and Systems*. Cham: Springer Nature Switzerland, pp. 85–97.

Achtibat, R., Hatefi, S.M.V., Dreyer, M., Jain, A., Wiegand, T., Lapuschkin, S. & Samek, W. (2024) *Attnlrp: Attention-aware layer-wise relevance propagation for transformers*. arXiv:2402.05602.

Acito, F. (2023) 'Ensemble models', in *Predictive Analytics with KNIME: Analytics for Citizen Data Scientists*. Cham: Springer Nature Switzerland, pp. 255–265.

Ahmed, M.W. & Jalal, A. (2024) 'Robust object recognition with genetic algorithm and composite saliency map', in *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*. IEEE, pp. 1–7.

Ahsan, M., Gomes, R., Chowdhury, M.M. & Nygard, K.E. (2021) 'Enhancing machine learning prediction in cybersecurity using dynamic feature selector', *Journal of Cybersecurity and Privacy*, 1(1), pp. 199–218.

Alanazi, F., Jambi, K., Eassa, F., Khemakhem, M., Basuhail, A. & Alsubhi, K. (2022) 'Ensemble deep learning models for mitigating DDoS attacks in software-defined networks', *Intelligent Automation & Soft Computing*, 33(2).

Al-Essa, M., Andresini, G., Appice, A. & Malerba, D. (2024) 'PANACEA: A neural model ensemble for cyber-threat detection', *Machine Learning*, pp. 1–44.

Alghanmi, N., Alotaibi, R. & Buhari, S.M. (2022) 'Machine learning approaches for anomaly detection in IoT: An overview and future research directions', *Wireless Personal Communications*, 122(3), pp. 2309–2324.

Ali, R., Ali, A., Iqbal, F., Hussain, M. & Ullah, F. (2022) 'Deep learning methods for malware and intrusion detection: A systematic literature review', *Security and Communication Networks*, 2022(1), p. 2959222.

Alserhani, F. & Aljared, A. (2023) 'Evaluating ensemble learning mechanisms for predicting advanced cyber attacks', *Applied Sciences*, 13(24), p. 13310.

Athief, R., Kishore, N. & Paranthaman, R.N. (2024) 'Web application firewall using machine learning', in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. IEEE, pp. 1–7.

Azam, Z., Islam, M.M. & Huda, M.N. (2023) 'Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree', *IEEE Access*, 11, pp. 80348–80391.

Betarte, G., Pardo, Á. & Martínez, R. (2018) 'Web application attacks detection using machine learning techniques', in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, pp. 1065–1072.
Brodie, M.L. (2024) *A framework for understanding data science*. arXiv:2403.00776.

Chen, C. et al. (2023) 'Application of GA-WELM model based on stratified cross-validation in intrusion detection', *Symmetry*, 15(9), p. 1719.

Chicco, D. & Jurman, G. (2020) 'The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation', *BMC Genomics*, 21(1), p. 6.

Choudhury, M.P. & Choudhury, J.P. (2022) 'Machine learning-based model to find out firewall decisions towards improving cyber defence', in *International Conference on Internet of Things and Connected Technologies*. Singapore: Springer Nature Singapore, pp. 179–195.

CrowdforThink (2019) *Bringing AI and machine learning accessible to enterprises: Credit to cloud*. Available at: https://crowdforthink.com/blogs/bringing-ai-and-machine-learning-accessible-to-enterprises-credit-to-cloud

Daram, K. & Senthilkumar, P. (2025) 'Optimizing Cloudflare security and performance with AI-based web application firewall and anomaly detection', *International Journal on Smart Sensing and Intelligent Systems*, 18(1). https://doi.org/10.2478/ijssis-2025-0040

Dawadi, B.R., Adhikari, B. & Srivastava, D.K. (2023) 'Deep learning technique-enabled web application firewall for the detection of web attacks', *Sensors*, 23(4), p. 2073.

Farha, F. & Ahmed, M.U. (2024) 'Heterogeneous ensemble approach in intrusion detection using stacking technique', in *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*. IEEE, pp. 1–6.

Franceschi, L. et al. (2024) *Hyperparameter optimisation in machine learning*. arXiv:2410.22854.

Gandomi, A.H., Chen, F. & Abualigah, L. (2022) 'Machine learning technologies for big data analytics', *Electronics*, 11(3), p. 421.

Gill, M.S. et al. (2023) 'Integration of domain expert-centric ontology design into the CRISP-DM for cyber-physical production systems', in *2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, pp. 1–8.

Javaid, A., Niyaz, Q., Sun, W. & Alam, M. (2016) 'A deep learning approach for network intrusion detection system', *EAI Endorsed Transactions on Security and Safety*, 3(9), p. 21.

Jeffrey, N., Tan, Q. & Villar, J.R. (2024) 'Using ensemble learning for anomaly detection in cyber–physical systems', *Electronics*, 13(7), p. 1391.

Kalariya, P., Jethva, M. & Alginahi, Y. (2024) 'ML assisted web application firewall', in *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE, pp. 1–6.

Kimanzi, R., Kimanga, P., Cherori, D. & Gikunda, P.K. (2024) *Deep learning algorithms used in intrusion detection systems—A review*. arXiv:2402.17020.

Kook, L. et al. (2022) *Deep interpretable ensembles*. arXiv:2205.12729.

Larner, A.J. (2024) 'Paired measures', in *The 2×2 Matrix: Contingency, Confusion and the Metrics of Binary Classification*. Cham: Springer International Publishing, pp. 17–53.

Li, M., Gao, Q. & Yu, T. (2023) 'Kappa statistic considerations in evaluating inter-rater reliability between two raters', *BMC Cancer*, 23(1), p. 799.

Li, W. et al. (2023) *Deep model fusion: A survey*. arXiv:2309.15698.

Li, Z. et al. (2023) 'Towards inference efficient deep ensemble learning', in *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(7), pp. 8711–8719.

Masum, M. et al. (2021) 'Bayesian hyperparameter optimization for deep neural network-based network intrusion detection', in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 5413–5419.

Mi, X., Zou, B., Zou, F. & Hu, J. (2021) 'Permutation-based identification of important biomarkers for complex diseases via machine learning models', *Nature Communications*, 12(1), p. 3008.

Mungoli, N. (2023) *Adaptive ensemble learning: Boosting model performance through intelligent feature fusion in deep neural networks*. arXiv:2304.02653.

Nassif, A.B., Talib, M.A., Nasir, Q. & Dakalbab, F.M. (2021) 'Machine learning for anomaly detection: A systematic review', *IEEE Access*, 9, pp. 78658–78700.

Nti, I.K., Quarcoo, J.A., Aning, J. & Fosu, G.K. (2022) 'A mini-review of machine learning in big data analytics: Applications, challenges, and prospects', *Big Data Mining and Analytics*, 5(2), pp. 81–97.

Odeh, A. & Taleb, A.A. (2024) 'Ensemble learning techniques against structured query language injection attacks', *Indonesian Journal of Electrical Engineering and Computer Science*, 35(2), pp. 1004–1012.

Ovi, M.S.I., Rahman, M.H. & Hossain, M.A. (2024) *PhishGuard: A multi-layered ensemble model for optimal phishing website detection*. arXiv:2409.19825.

Owusu-Adjei, M. et al. (2023) 'Imbalanced class distribution and performance evaluation metrics', *PLOS Digital Health*, 2(11), p. e0000290.

Panasov, V.L. & Nechitaylo, N.M. (2021) 'Decision trees-based anomaly detection in computer assessment results', *Journal of Physics: Conference Series*, 2001(1), p. 012033.

Pansari, N. et al. (2024) 'Attack classification using machine learning on UNSW-NB15 dataset', in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. IEEE, pp. 1–9.

Putra, Z.P. (2024) 'Evaluating the performance of classification algorithms on the UNSW-NB15 dataset for network intrusion detection', *Jurnal Ilmiah FIFO*, 16(1), p. 84.

Saleem, M.A. et al. (2022) 'Comparative analysis of recent architecture of convolutional neural network', *Mathematical Problems in Engineering*, 2022(1), p. 7313612.

Sambandam, R.K. et al. (2023) 'Comparison of machine learning-based intrusion detection systems using UNSW-NB15 dataset', in *International Conference on Artificial Intelligence on Textile and Apparel*. Singapore: Springer Nature Singapore, pp. 311–324.

Sissodia, R. et al. (2025) ‘Artificial intelligence (AI) in cybersecurity’, in *Advances in Computational Intelligence and Robotics*. Hershey, PA: IGI Global, pp. 121–152.

Stormit.cloud (2022) *StormIT achieves AWS service delivery for AWS WAF*. Available at: https://www.stormit.cloud/blog/aws-waf-service-delivery/

Tama, B.A., Nkenyereye, L., Islam, S.R. & Kwak, K.S. (2020) ‘An enhanced anomaly detection in web traffic using a stack of classifier ensemble’, *IEEE Access*, 8, pp. 24120–24134.

Tan, P. (2023) *Ensemble-based hybrid optimization of Bayesian neural networks and traditional machine learning algorithms*. arXiv:2310.05456.

Tian, Z. et al. (2019) ‘A distributed deep learning system for web attack detection on edge devices’, *IEEE Transactions on Industrial Informatics*, 16(3), pp. 1963–1971.

Vibhute, A.D. et al. (2024) ‘Network anomaly detection and performance evaluation of convolutional neural networks on UNSW-NB15 dataset’, *Procedia Computer Science*, 235, pp. 2227–2236.

Waheed, M. et al. (2023) *An evaluation and ranking of different voting schemes for improved visual place recognition*. arXiv:2305.05705.

Wang, Y. & Wang, X. (2022) *A unified study of machine learning explanation evaluation metrics*. arXiv:2203.14265.

Wang, Z., Ouyang, Y. & Zeng, H. (2024) ‘ARFN: An attention-based recurrent fuzzy network for EEG mental workload assessment’, *IEEE Transactions on Instrumentation and Measurement*.

Wirth, R. & Hipp, J. (2000) ‘CRISP-DM: Towards a standard process model for data mining’, in *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 1, pp. 29–39.

Xu, M. et al. (2024) *A survey of resource-efficient LLM and multimodal foundation models*. arXiv:2401.08092.

Yang, F. et al. (2023) ‘Assessing inter-annotator agreement for medical image segmentation’, *IEEE Access*, 11, pp. 21300–21312.

Yang, Y., Lv, H. & Chen, N. (2023) ‘A survey on ensemble learning under the era of deep learning’, *Artificial Intelligence Review*, 56(6), pp. 5545–5589.

Zhang, X.Y. & Wang, M.M. (2023) ‘An efficient combination strategy for hybrid quantum ensemble classifier’, *International Journal of Quantum Information*, 21(06), p. 2350027.

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 104-116
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1737

Mission Statement Attributes and Employee Engagement in the Nigerian Banking Sector: Evidence from Ogun state, Nigeria

[1]Banjo Hassan, [2]Victor Adeyeye,
[3]Hassanat Hassan & [4]Kehinde Shosanya
[1, 2, 3 & 4]Olabisi Onabanjo University

Abstract

Nigerian banks have mission statements like many other corporate firms; they struggle to keep their workforce engaged. One critical yet often overlooked factor is the clarity and impact of the mission statement. Literature has not documented the nexus between attributes of these mission statements and employee engagement. As a guiding principle, a mission statement has the potential to inspire and align employees with organizational goals and values. Against transformational leadership theory, this study examines the impact of mission statement on employee engagement in the banking sector in Ogun State. The study employed the descriptive survey design. A sample of 154 employees was selected from nine deposit money banks in Ijebu-Ode, Ogun state for the study. Primary data was collected using a well-validated instrument. The findings from the regression analysis revealed that clarity and specificity of mission statement has positive significant effect on employee engagement. Also, effective communication of mission statement has positive significant effect on employee engagement. Moreover, employee personal connection to mission statement has positive significant effect on employee engagement. The three independent constructs also collectively have significantly predict employee engagement. The study concluded that mission clarity, its effective communication, and employees' personal connection to it significantly boosts employee engagement. . Organizations aiming to improve engagement should prioritize not only the content of their mission but also how it is shared and internalized, ensuring it resonates personally with employees and drives a shared sense of purpose and direction.

Keywords: Employee Engagement; Mission statement; Clarity/specificity of mission statement; Personal connection to mission statement; Effective communication.
Wordcount: 239

1.0 Introduction

Globally, reports have indicated a drop in the level of employee engagement in organisations. Gallup Institute reported that 85% of employees are disengaged in their day-to-day routine (Gallup, 2017). This according to Motyka (2018) is due to the absence of support in terms of organizational direction and road map that allows employees achieve what matters to them. In the Nigerian banking sector, low employee engagement manifests in high and increasing staff turnover. The Central Bank of Nigeria (2021) shows that turnover rate in the banking sector jumped from 12.5% in 2018 to 28.1% in 2021 caused by poor work engagement among others Obazi et al (2023). This pushes the concreteness of mission statement in promoting not only organizational performance but also employee engagement, which represents a little niche in improving the overall corporate productivity (Burhan and Khan 2024).

Organisations are basically utility creating economic unit of the society. The efficiency with which they create utility by converting the resources into need satisfying products (goods and services) remain the focus of every management. This often depend on the enthusiasm, dedication and commitment brought to work by the employees. An engaged employee show enthusiasm and dedication to work and exhibit concern about performance of the company (Kurniawati & Makhmut, 2023). Management must therefore implement strategies that can reduce employees' disengagement (Haris & Yang, 2023). An organization business strategy and performance, to some extent depends on the mission statement and core values of the organization. Mission statement is purposeful in building communication internally to employee and externally to stakeholders to achieve organizational goals. The mission statement inscribes an organization culture, values and operational floes that makes the organization unique among its other competitors (Khripunov, 2023; Dermol and Širca 2018; Cortés-Sánchez, et al., 2019). However, apart from the attributes of mission statement to spur employee direction and engagement in an organization, it is a core component in the strategic management of an organization (Deasy et al, 2023).

Intrinsically, mission statement is an important tool to motivate and engage employees because it guides employee behaviour and informs how he interacts with customers. The development of accurate and purpose driven mission statement becomes relevant in ensuring employee motivation and productivity of the organization at large (Akhtar et al., 2016; Gede and Huluka, 2024). A well-articulated mission statement helps to shape employee's pattern of behaviour, increase their level of commitment and spur performance. The continuity prospects and goodwill of an organization largely depend on the mission statement because it shows the direction where the organization is headed in the future. It enables the employee to integrate themselves into the moving wheel of the organization.

Employee engagement is a key driver of organizational success, influencing productivity, job satisfaction, and retention (Analoui and Karami., 2002; Gede and Huluka, 2024). However, many organizations struggle to keep their workforce engaged. One critical yet often overlooked factor is the clarity and impact of the mission statement. As a guiding principle, a mission statement has the potential to inspire and align employees with organizational goals and values. However, its true influence on engagement remains underexplored. While organizations carefully craft mission statements, their impact depends on effective communication and reinforcement an aspect that requires deeper analysis (Cortés-Sánchez and Rivera, 2019; Jonyo, et al 2018).

Nigerian banking sector continues struggle with high employee turnover rates and low engagement levels. The Central Bank of Nigeria report (2021) for example confirmed over 35% increase in employee turnover within a single year. Although, most of the Nigerian banks have mission statements, the effect of their clarity, compelling nature or effective communication to their workforce engagement (Satata, 2021; Susanto, et al., 2023) remain unstudied. This oversight can lead to misalignment between the organization's objectives and employees' personal values, resulting in diminished motivation and commitment. Additionally, the cultural and socio-economic factors unique to Nigeria may further complicate the relationship between mission statements and employee engagement. (Satata, 2021; Susanto, et al., 2023).

In the Nigerian context, research has not adequately captured the construct of mission statements from the perspective of their channelling attributes within an organization or their role in shaping employee attributes. Specifically, there is a lack of studies addressing how mission statements inform employees and foster employee involvement and connection in their development. Additionally, research on the relationship between mission statements and employee engagement in the Nigerian banking sector remains scanty. This study aims to bridge this gap by contributing to the body of knowledge on the subject. Specifically, we seek to answer the following research questions: what is the influence of Clarity/specificity of mission statement on employee engagement in the Nigerian banking sector; Does effective communication of mission statement impact on employee engagement in the Nigerian banking sector and How does employee personal connection to mission statement impact on their engagement Employee personal connection to mission statement has no significant impact on employee engagement?

The remainder of this paper is organized as follows: review of relevant literature and methodology. Next, the results are outlined and discussed, followed by conclusions and recommendations

2.0. Literature Review

2.1. Employee Engagement

Employee work engagement was described as a positive state of vigour, dedication, and absorption in a task (Schaufeli et al., 2006). This definition suggest that engagement is behavioural, marked by employee emersion in their tasks. Attempting to summarise different earlier definitions, McKenzie (2025) defined employee engagement as the degree of employees' commitment to their organizational goals as demonstrated by their thoughts, feelings, actions, and their emotional connection towards their organization, their work, and their team. Bakker and Woerkom (2017) definition of employee engagement agrees that it is an emotional and multidimensional psychological construct denoting employees' emotional ties with and deep passion for the job and organization. Because it is attitudinal, specifically, about employees' perceived personal evaluation of their emotional and social attachment to their work and its environment, Byrne et al, (2016) noted that employee engagement is inherently difficult to gauge and more importantly, improve. Notwithstanding, what matters to managers is the manifestation of attitudes that points to engagement in workplace behaviours. Essentially, Schaufeli and Bakker (2004) concluded that employees are engaged when they exhibit energy, investment of effort, and persistence in the face of challenges; sense of significance, passion, and pride in their work; and full attention and to their tasks.

Since its introduction in 1990 by William Kahn, the concept has been a widely studied behaviour in organizational studies and management (Srividya, 2018). Kahn (1990) asserted that engagement manifests when individuals are deeply involved in their work, exert effort, focus, and commitment to advancement of tasks. Gupta and Sharma (2016), in their study, explored the meaning and key drivers of employee engagement through a review of existing literature. They shared various definitions, including one that describes engagement as a sustained, positive, motivated state of fulfillment marked by high energy and enjoyment (Maslach et al., 2001), and another that views it as an employee's mental, emotional, and behavioural focus on achieving organizational goals (Shuck & Wollard, 2010). For the purposes of this study, employee engagement is seen as employees' passionate and enthusiastic involvement in working toward the organization's goals and mission.

Three main dimensions of engagement have been identified in the literature: intellectual, affective and social. Intellectual engagement involves employees thinking carefully about their jobs and finding ways to improve them. Affective engagement is seen in employees who feel positively about their work while social engagement refers to employees actively sharing ideas and collaborating with colleagues to enhance work processes. Macey and Schneider (2008) noted that engagement that includes organizational commitment (how strongly employees feel attached or loyal to the organization), job involvement (which covers both dedication to tasks and commitment to one's role), and empowerment (which depends on an employee's confidence in their ability to do their work well). They also emphasized engagement behaviours, which go beyond basic job expectations and reflect an employee's desire to help the organization grow.

Kahn (1990) classical classification of employee engagement suggest three different dimensions: cognitive, emotional, and behavioural, which leads to positive organizational outcomes (Saks, 2006; Shuck & Wollard, 2010). In support of these dimensions, Shuck et al. (2017: 269) defined employee engagement as the "positive, active, work-related psychological state operationalized by the maintenance, intensity, and direction of cognitive, emotional, and behavioural energy".

Cognitive engagement, according to Shuck (2020), refers to employees' evaluation of their work environment and tasks. It occurs when employees feel "that their work mattered, that they were supported in their work, and that their well-being was considered fairly" (Shuck et al., 2014: 245). Accordingly, a cognitively engaged employee shows mental and psychological commitment to task at hand, keep intensity and move in the right direction until task completion (Shuck, 2020). Emotional engagement concerns individual's emotional connection with his organisation, leading him to contribute intangible personal resources like knowledge and pride (Shuck & Reio, 2014). It is the positive emotion that stem from perceived organizational support (Rich et al., 2010). The positive emotions of pride and trust come from employees' subjective evaluations about work environment during a cognitive engagement. Behavioural engagement is the physical manifestation of cognitive and emotional engagement. It refers to higher levels of extra-role behaviour which results from cognitive and emotional engagement and is usually characterized by putting in more effort toward organizational goals (Barnes et al., 2014).

Macey and Schneider concluded that when all these elements come together, they create genuinely engaged employees. Such engagement is not just beneficial but offers organizations a clear competitive advantage. Supporting this idea, Harter et al. (2002) found in their review that employee engagement and satisfaction are linked to better outcomes like increased customer satisfaction, higher productivity, greater profitability, stronger employee retention, and improved workplace safety. Employee disengagement on the other hand causes low commitment, high absenteeism, high turnover intention (V. Gupta & Kumar, 2013; Macey et al., 2009; Mone & London, 2009). Highly engaged employees report high extra-role behaviours, are less often absent, show high commitment and job satisfaction (Agarwal, 2014; Saks, 2006).

2.2. Mission Statement

Mission statements are an integral part of corporate strategy, defining an organization's purpose and guiding employees to work both independently and collectively toward achieving company goals (Kotler et al., 2008). As a fundamental element of strategic planning, mission statements have been studied since the early 1970s, with Drucker (1971) describing them as a tool for setting clear and realistic business objectives. An organization's mission defines its reason for existence, while a mission statement is a formalized document that captures its purpose and direction (Desmidt and Prinzle, 2007). It aligns organizational processes with business strategies and serves as a foundation for consistency in purpose. A mission statement addresses key stakeholder questions, such as why the organization exists, its objectives, and what it aims to achieve. It also outlines the organization's scope, distinguishing it from competitors by highlighting its products, services, markets, and technology (Inyang, 2004).

A mission statement that inspires employees is effective (Ganu, 2013). It fosters shared expectations among employees and communicates the firm's public image to stakeholders (Analoui & Karami, 2002). It sets boundaries for resource allocation and strategic decisions, becoming meaningful only when it influences behaviour and guides actions (Sufi & Lyons, 2003). According to Bart (1997), having a mission statement engages and motivates organizational members as well as facilitates consistent and focused resource allocation. Ezekwe and Egwu (2016) contend that apart from indicating present and future direction of a company, mission statements motivate performance. Malbašić, Rey, and Posarić (2018) concluded that aligning personal missions with organizational missions motivates employees.

Many corporations emphasize mission statements due to their role in helping businesses meet their goals. Despite their importance, some entrepreneurs overlook them, making it harder to articulate their business vision and objectives. Employees on the basis of mission statement can decide if they are willing to expend effort towards the mission if it is motivating, engaging and resonates with them. Klemm, Sanderson and Luffman (1991) also agree mission statement perform dual roles: enhance organization's image externally and motivate employees.

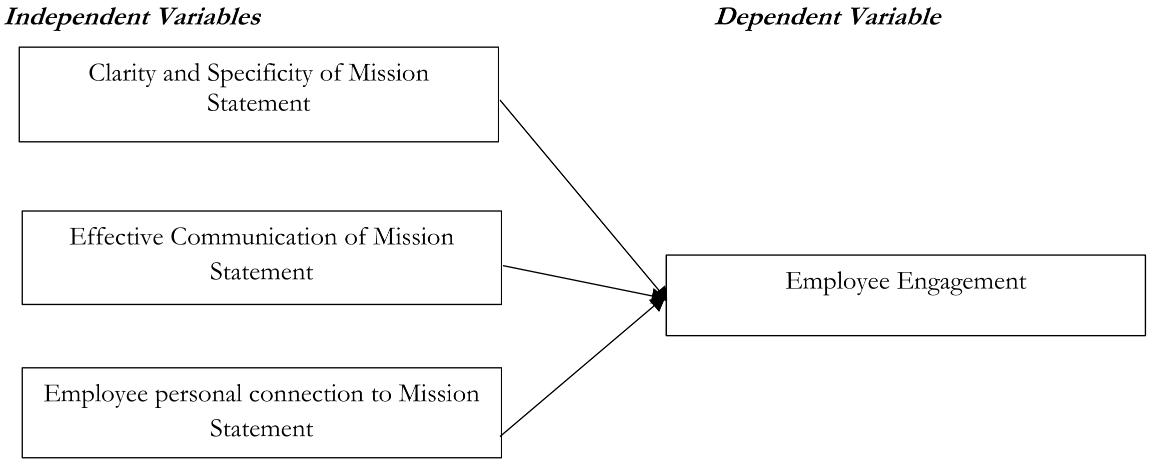

Figure 1: Conceptual Framework
Author's Design, 2025

2.3. Theoretical Review

Transformational Leadership Theory

This concept, initially introduced by James MacGregor Burns and later expanded by Bernard M. Bass, highlights the ability of leaders to inspire and energize their followers toward achieving mutual goals (Adigwe, 2024). It presents a clear vision, promotes creativity, and supports personal development. In this study, the theory offers valuable perspectives on managers' roles in shaping organizational direction. It has been noted by Avolio, Walumbwa, and Weber (2009) that leaders can achieve shared goals by effectively communicating visions and missions. In addition, the theory emphasizes the need to harmonize organizational components with individual principles to build a unified culture. (Bass and Riggio, 2006; Avolio and Yammarino, 2013). Accordingly, Higgs and Rowland (2005) noted that such strategic alignment ensures focus, direction, and consistency in organizational efforts by articulating priorities clearly. In short, Transformational Leadership Theory underscores the understanding of the role of managers in stimulating collective action to facilitate organizational progress.

Signalling Theory

Signalling theory suggests that organizations can enhance their reputation and attract investors by transparently sharing information about their competitive advantages with the public. According to Klemm et al. (1991), one of the key roles of a mission statement is to strengthen an organization's image while also serving as a motivational tool for employees. As a result, it is believed that high-performing organizations tend to present more detailed mission statements in their annual reports and through other avenues. Regarding the function of a mission statement within an organization, Klemm et al. (1991) argue that its primary purpose is to engage with external stakeholders, with employee motivation being a secondary objective. However, later findings revealed that mission statements can serve additional purposes beyond these initial assumptions.

2.4. Empirical Review

David (2020) examined the effect of mission statement that is peculiar to the attributes of fortune 500 in relations to their organizational performance and reported that mission and vision statements of the fortune 500 have positive significant effect on non-financial measure of performance and no significant impact on financial performance. Although with no theoretical framework, Sebastian (2016) examined the perceived quality of mission statement from four perspectives which includes mission comprehension ambiguity, behavioral integrity, self-efficiency and person-organization fit on employee mission engagement. Analysing responses from 1,418 employees from Belgian public social service organization that had just readjusted its mission statement using univariate and bivariate analysis, the study reported among others that perceived message quality is a correlate of employee engagement.

In a more recent study, Kipasika (2024) had hinged his study on organizational culture theory, transformational leadership theory, and servant leadership theory to analyse the relationship between staff categories, understanding and recognition of corporate mission, among others in Tanzania. The study revealed significant differences in the correct identification rates of mission across staff categories, signifying need for differentiated target communication and engagement strategies. More so, staff position did not predict awareness of organizational mission. However, Az-Zaakiyyah, et al (2024) study of mediating role of vision and mission in the relationship between corporate culture and employee performance reported that clear vision and mission significantly influence employee performance.

Deasy, et al (2023) employed survey using purposive sampling to examine if self-efficacy and internal marketing and employee engagement influence attainment of Vision and Mission. Findings show that employee engagement acts as a partial mediating variable between self-efficacy and achievement of the Vision and Mission. On their part Dobrinić and Fabac (2021) study focused on testing if organizational commitment and job satisfaction of employee differs based on their knowledge of the organizations mission and vision. T-test analyses of primary data obtained indicated that there exist a statistically significant differences in job satisfaction between employees who are acquainted with the mission of their organization and those who are not and that this difference is especially noticeable among employees who work for the government or public organizations.

Similarly, Dermol and Širca (2018) reported that that there exists a positive connection between the existences and communication of company mission and performance of that organization. Kuye and Ezebuiro (2023) revealed that mission statement has positive significant effect on employee motivation, employee commitment, and company direction while Taiwo and Lawal (2016) conclusion was similar to the findings of Kuye and Ezebuiro (2023) when they noted that carefully crafted and implemented mission statements is capable of influencing employee in their day-to-day tasks and promote the achievement of organizational goals.

3.0. Methodology

This study adopted a descriptive survey design, a method used to assess and quantify the characteristics of a specific variable or phenomenon. This approach was chosen because it enables the collection of structured data from a target population, providing a clear and measurable understanding of the subject matter. The descriptive survey design is particularly effective in capturing respondents' perceptions, behaviours, and attitudes, making it well-suited for this study. The population of this study captures employees of the deposit money banks operating in Ijebu-ode, the commercial centre of Ijebu division, Ogun state. They include Access Bank, First Bank, United Bank for Africa, Guaranty Trust Bank, First City Monument Bank, Wema Bank, Sterling Bank and Polaris Bank.

Table 1: Study Population Distribution

Bank	Senior staff	Junior Staff	Total
FBN	7	10	17
GTB	6	14	20
Access	7	8	15
UBA	6	12	18
Zenith	5	11	16
FCMB	9	17	26
Wema	6	7	13
Sterling	5	12	17
Polaris	4	9	13
TOTAL	**55**	**99**	**154**

Source: Field Survey

The study employed a stratified sampling technique where population was divided into the two subgroups of the population and samples were selected proportionately. The sample size of the study comprises of the census population of the nine banks branches in Ijebu-ode. Primary data used in the study was sourced from the respondents through a well-structured questionnaire. The dependent variable (Job engagement) was measured by Houle et al (2022), a well validated and reliable scale widely used in the measurement of employee engagement while mission statement (Clarity and Specificity; Effective Communication; personal connection) was self-constructed.

The independent variables and dependent captures 21 items in total with a close-ended questions. The instrument on mission statement was validated the using content validity (CVI) test by first exposing the instrument to colleagues. All the modified questionnaires were returned to the evaluators for final evaluation on a two-point scale of Relevant (R) and Irrelevant (I) after careful implementation of suggested recommendations. We then conducted content validity index (CVI) with a value of 0.753. The internal consistency of the measurement scales was assessed using Cronbach's alpha with $\alpha= 0.73$. More so, common method bias was tested using Harman's single-factor test. All measurement items were entered t the same time into an unrotated exploratory factor analysis using principal component analysis. The result showed that multiple factors emerged from the analysis. The first factor accounted for (38.6%) of the total variance (below 50% threshold recommended in the literature (Podsakoff et al., 2003) indicating no common method bias. The collected data was analysed using the descriptive (mean, standard deviation and percentages) and linear regression.

3.2. Model Specification

The simple form of the model for the study is given below;

$EE_{EE}=f(MS_t)$……………………………………………………….1

$MS_t=f(CSM_{(t,)}\ ECM_{(t,)}\ EPC_{(t,)})$………………………..2

The models include the following;

Objective one model

Model 3 below captures the impact of clarity/specificity of mission statement, on employee engagement in the banking sector.

$EE_t= (\alpha_0+ CSM_t+\mu_t)$……………….3

Model for objective two

The model 4 presented below captures the impact of effective communication of mission statement on employee engagement in the banking sector in Ogun state.

$EE_t= (\alpha_0+ ECM_t+\mu_t)$……………….4

Model for objective three

Model 5 presented below captures the effect of employee personal connection to mission statement on employee engagement in the banking sector in Ogun state.

$EE_t= (\alpha_0+ EPC_t+\mu_t)$……………….5

Model for objective four

The model 6 presented below captures the impact of clarity of mission statement, effective communication of mission statement, employee personal connection to mission statement on employee engagement in the banking sector in Ogun state.

$EE_t= (\alpha_0+ CSM_t+ ECM_t+ EPC_t+\mu_t)$……………….6

Where:
EE; Employee Engagement at time t
MS; Mission statement at time t

Dependent Variable
EE; Employee Engagement at time t

Independent Variable
CSM: Clarity and Specificity of Mission Statement at time t
ECM: Effective Communication of Mission Statement at time t
EPC: Employee personal connection to Mission Statement at time t

4.0. Results and Discussion

4.1. Results

Table 2: Demographic attributes of respondents

Demographic	Frequency	Percentage
Sex		
Male	73	47.4
Female	81	52.6
Total	154	100
Age		
20-25	14	9.4
26-34	88	57.8
35-45	39	25.0
46 and above	13	7.8
Total	154	100
Level of Management		
Junior	99	64.3
Senior	55	35.7
Total	154	100
Years of Experience		
1-5years	14	9.3
6-10years	41	26.6
11-20years	99	64.1
Total	154	100

Source: Researcher's Field Survey, 2025

The table 2 presents the demographic analysis of the gender of 154 respondents. While 47.4% of the respondents were male, 52.6% of the respondents were female. Respondents within the ages of 26-34 representing 57.8% form the largest age group. The respondents are also majorly comprised of senior level employees who are expected to be more alive to the mission statement of their firms and exhibit higher engagement as studies have shown that both age and job tenure are correlates of commitment as work engagement increases with age.

Table 3: Descriptive Analysis of variables

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Employee_Engagement	154	5.00	13.00	5.3167	2.44604
Clarity_Specificity_MS	154	4.00	10.00	3.7256	2.14359
Effective_Communicatio_MS	154	4.00	12.00	3.0625	2.09970
Employee_Personal_con_MS	154	4.00	12.00	3.1094	2.26160
Valid N (listwise)	154				

Author's Computation, 2025

Table 3 showed *employee engagement* has a mean value of 5.3167 with the highest value of 13.00 and lowest value of 5.00 with standard deviation of 2.44 shows that level of dispersion of the responses from the mean value is 2.44. *Clarity and specificity* of mission statement has a mean value of 3.7256., the highest value was 10.00 and lowest value of 4.00 while the standard deviation shows of 2.14 lower than employee engagement shows the responses were more consistent. Effective communication of mission statement has a mean value of 3.0625 and standard deviation shows that level of disparity of the group from the mean value is 2.09. Employee personal connection to mission statement has a mean value of 3.1094with standard deviation of 2.64. The statistics show that while employee engagement is moderately high, employee personal connection to the mission statements and communication are both moderately low.

Table 4: Correlation Matrix

Correlations

	Engagement	Clarity/Specificity	Effective Communication	Personal connection
Employee Engagement	1	.464**	.309**	.365**
Clarity/Specificity	.464**	1	.457**	.685**
Effective_Communicatio_MS	.509**	.357**	1	.234**
Employee_Personal_con_MS	.465**	.485**	.534**	1

** Correlation is significant at the 0.01 level (2-tailed).

Author's Computation, 2025

The correlation table above shows that there is no multicollinearity among the constructs of the independent variables although all of them have moderate correlations with the dependent variable, employee engagement.

Test of Hypothesis

Table 5: H$_{0=}$ Clarity/specificity of mission statement has no significant impact on employee engagement

Variable	Co-efficient	Std-Error	t-stat	P-value
Constant	1.551	0.518	2.994	0.004
Clarity_Specificity	0.864	0.073	13.489	0.000
R²	0.746	**F.cal**	181.951	
Adj. R²	0.742	**Sig.F**	0.000	

Source: Author's Computation, 2025 Dependent Variable: Employee Engagement

Table depicted that clarity and specificity of mission statement has positive significant effect on employee engagement at *(β_1= 0.864; ϱ<0.05).* The {F-cal= 181.951, ϱ< 0.05}, showed that the overall model is statistically significant at 5% level of significance. The R^2 value of 0.746 is a measure of goodness of fit of the regression model and indicated that 74.6% of variation in employee engagement can be explained by how clear and specific the mission statement of their banks are. Hence, if mission statement's clarity is increased by a single unit of measure, it can result in improvement of employees' engagement by 74.6%.

Table 6: H$_{0=}$ Effective communication of mission statement has no significant impact on employee engagement.

Variable	Co-efficient	Std-Error	t-stat	P-value
Constant	1.799	0.395	2.994	0.000
Effective communication	0.585	0.53467	17.174	0.000
R²	0.342	**F.cal**	294.948	
Adj. R²	0.329	**Sig.F**	0.000	

Source: Author's Computation, 2025 Dependent Variable: Employee Engagement

Table 6 presents the test of hypotheses on the impact of effective communication of mission statement on employee engagement of bank staff. The analyses show a positive significant effect of effective communication of mission statement on employee engagement at *(β_1= 0.585; ϱ<0.05).* The table revealed {F-cal= 294.948, ϱ< 0.05}. The F-test tests the null hypothesis that all the slope coefficients (safe the intercept) are zero and that the independent variable has no explanatory power on the dependent variable. Since the F-statistic is large and its associated p-value is small less than 0.05, we reject the null hypothesis, to the effect that the predictor is significantly related to the dependent variable. There is enough evidence to reject the null hypothesis that *"effective communication of mission statement has no relationship with engagement"*. Thus, communication of mission statement explains variation in employee engagement. Meanwhile, R^2 is a measure of goodness of fit of the regression model. It revealed that, the independent variable effective communication of mission statement for 0.826 explains 82.6% variation or change in the dependent variable- employee engagement.

Table 7: H$_{0=}$ Employee personal connection to mission statement has no significant impact on employee engagement.

Variable	Co-efficient	Std-Error	t-stat	P-value
Constant	0.795	0.267	2.980	0.004
Employee_personal	0. 546	0.4436	29.167	0.000
R²	0.298	**F.cal**	850.696	
Adj. R²	0.272	**Sig.F**	0.000	

Author's Computation, 2025 Dependent Variable: Employee Engagement

Table 7 show the results of the test of hypothesis *Employee personal connection to mission statement has no significant impact on employee engagement*. The F-calculated value is **850.696** and its associated p-value is less than 0.05. The null hypothesis for the overall regression model is that *the slope of the independent variable = 0* (i.e. the model has no explanatory power). But because $p < 0.05$, we reject the null hypothesis. In other words, the model **explains a statistically significant portion of variance** in employee engagement better. Thus, the regression is statistically significant overall. The coefficient $\beta_1 = 0.546$ means that for every one-unit increase in *employee personal connection to mission statement*, *employee engagement* is expected to increase by 0.96.5 unit More so, because the t-statistic is 29.167 and its p-value is < 0.05, the coefficient is statistically significant. That is, we reject the null hypothesis that $\beta_1 = 0$ (i.e. "no effect"). The very large t-statistic also suggests a strong signal to this conclusion as the coefficient is far from zero relative to its standard error). Hence, employee personal connection to mission statement has a positive and significant effect on employee engagement.

Table 8: $H_{0=}$ Clarity/specificity of mission statement, effective communication of mission statement and employee personal connection to mission statement jointly do not significant impact on employee engagement.

Variable	Co-efficient	Std-Error	t-stat	P-value
Constant	0.387	0.181	2.139	0.036
Effective_communication	0.242	0.056	5.063	0.000
Clarity_Specificitiy	0.149	0.048	3.503	0.001
Employee_personal connection	0.647	0.043	16.281	0.000
R^2	0.147	**F.cal**	20.982	
Adj. R^2	0.145	**Sig.F**	0.000	

Author's Computation, 2025 Dependent Variable: Employee Engagement

Because the three predictors is most cases are likely to exist and influence engagement concurrently, a multiple regression was carried out to examine the effect of Effective Communication, Clarity/Specificity, and Employee Personal Connection to the Mission Statement on employee engagement. The Coefficient of the constant (intercept) = 0.387, p = 0.036 (< 0.05) shows the predicted value of employee engagement when all predictors (Effective Communication, clarity/Specificity, and Employee Personal Connection are zero. The **Coefficients of predictors** show the *marginal effect* of a one-unit increase in each independent variable on the dependent variable, holding all other variables constant. While **Effective communication** has a coefficient of 0.242 (a unit increase predicts **0.242 unit** increase in employee engagement). The t= 5.063; p = 0.000 indicate this is highly statistically significant; **Clarity/Specificity of mission statement has a** coefficient of 0.149 indicating a unit increase is associated with a **0.149 unit** increase in engagement. This is also significant given the t = 3.503; p = 0.001. Conclusively, **Employee personal connection to mission statement showed a** coefficient of 0.647. This is the strongest of the three effects with a one-unit increase predicting a **0.647 unit** increase in engagement. This is also very strongly significant.

The regression model was statistically significant, F-cal = 20.982, $p < 0.05$, and accounted for 14.7 % of the variance in engagement ($R^2 = 0.147$, Adjusted $R^2 = 0.145$). All three predictors had significant positive effects on engagement: Effective communication ($\beta = 0.242$, $t = 5.063$, $p < 0.001$); Clarity/Specificity ($\beta = 0.149$, $t = 3.503$, $p = 0.001$); Employee personal connection to mission ($\beta = 0.647$, $t = 16.281$, $p < 0.001$). These results suggest that while all three dimensions significantly influence engagement, employee personal connection has the strongest unique effect (i.e., the largest coefficient), controlling for the other variables.

4.2. Discussion of Findings

The findings depicts that clarity and specificity of mission statement has positive significant effect on employee engagement. Also, effective communication of mission statement has positive significant effect on employee engagement. This implies that organizations aiming to boost employee motivation and alignment with corporate goals should prioritize crafting mission statements that are both precise and meaningful. Moreover, consistent communication of these statements across all organizational levels can foster a deeper sense of purpose among employees, leading to increased commitment and productivity. Leaders and managers should therefore invest in strategies that not only define the mission clearly but also ensure it is regularly and effectively conveyed to all employees for maximum engagement. The findings agree with the works of David (2020) and Sebastian (2016).

The findings depicts that employee personal connection to mission statement has positive significant effect on employee engagement. This implies that when employees feel emotionally connected to the mission and are actively involved in shaping or living it out, their level of commitment and enthusiasm increases. Organizations should therefore foster a culture where employees understand how their roles align with the mission and are encouraged to participate in mission-related initiatives. Such involvement not only strengthens the sense of belonging but also boosts morale, motivation, and overall organizational performance through heightened engagement. The findings agree with the works of Kipasika (2024) and Az-Zaakiyyah, et al (2024).

The findings reveal effective communication of mission statement, clarity and specificity of mission statement and employee personal connection have positive significant effect on employee engagement. The findings imply that organizations seeking to boost employee engagement should prioritize crafting clear, specific, and meaningful mission statements. Effective communication of the mission ensures employees understand organizational goals, fostering alignment and motivation. When the mission resonates personally with employees, it strengthens their emotional commitment, leading to higher engagement, productivity, and retention. These results highlight that mission statements are not merely formalities but strategic tools that, if well-designed and well-communicated, can enhance organizational culture and performance by deeply connecting employees to the company's purpose. The findings agree with the works of Abin, et al (2024) and Deasy, et al (2023)

5.0. Conclusions And Recommendations

This study examines into the impact of mission statement on employee engagement of selected deposit money banks in Ijebu-Ode, Ogun State. The study analysed the impact of clarity/specificity of mission statement on employee engagement; the effect of effective communication of the mission statement on employee engagement; the effect of employee personal connection to mission statement on employee engagement; and the combined effect of the three on employee engagement.

The findings from this study demonstrate that clarity and specificity of the organizational mission statement, effective communication of the mission, and employees' personal connection to it each independently and collectively predict higher levels of employee engagement. Specifically, when the mission statement is clear, specific, and well-communicated, employees better understand and align their daily efforts with organizational goals, resulting in greater involvement, enthusiasm, and dedication at work. Furthermore, employees who develop a genuine personal connection to the mission, feeling that it resonates with their own values and sense of purpose exhibit even stronger engagement, as this emotional and participatory bond fosters deeper commitment and motivation.

In combination, these attributes of a mission statement create a powerful synergy that can significantly enhances overall employee engagement. This stresses the critical role of a thoughtfully crafted mission statement not only as a strategic tool but also as a foundation for building a motivated and committed workforce.

This study contributes to the literature by linking key attributes of mission statements (clarity/specificity, effective communication, and personal connection) to employee engagement through the lens of transformational leadership theory (Bass, 1985; Bass & Riggio, 2006). Transformational leaders inspire followers by articulating a compelling, value-aligned vision, using inspirational motivation to communicate organizational goals clearly and enthusiastically, and fostering individualized consideration that helps employees internalize and connect personally with the mission. The present findings extend this framework by showing that well-defined and effectively communicated mission statements serve as a key mechanism through which transformational leadership behaviours elevate followers beyond self-interest, heighten awareness of collective purpose, and promote engagement. The study recommends that HR leaders and corporate managers in the banking sector should craft clear and specific mission statements, avoid vague or overly broad language and instead, develop concise statements that articulate purpose, values, and direction in understandable terms. They should also regularly share the mission especially through digital communications with staff to reinforce their vigour.

References

Abdul-Azeez, O., Ihechere, A. O., & Idemudia, C. (2024). Transformational leadership in SMEs: Driving innovation, employee engagement, and business success. *World Journal of Advanced Research and Reviews*, *22*(3), 1894-1905.

Abin, G., Llorca, R., Sale, S. G., & Ramirez, C. S. (2024). The Power of Vision and Mission Statements: A Study on Their Impact on Micro Business Enterprises in Camarines Norte, Philippines. *Jurnal Genesis Indonesia*, *3*(2), 100-112.

Abun, D., Martin, K. O., Cabillo, E. R., Encarnacion, M. J., & Magallanes, T. (2022). Humanistic management and the counterproductive behaviours as perceived by the employees. *International Journal of Business Ecosystem & Strategy*, 4, 37–47. https://dx.doi.org/10.36096/ijbes.v4i3.348

Adigwe, C. S. (2024). Transformational Leadership: A Comparative Exploration of the Leadership Prowess of Jeff Bezos and Steve Jobs. *Asian Journal of Economics, Business and Accounting*, 24(3), 68-89

Akhtar, A., Nawaz, M.K., Mahmood, Z., Shahid, M.S. (2016). Impact of high-performance work practices on employees' performance in Pakistan: Examining the mediating role of employee engagement. *Pakistan Journal of Commerce Social Sciences*, 10(3), 708–724.

Analoui, F., & Karami, A. (2002). CEOs and development of the meaningful mission statement. *Corporate Governance*, *2*(3), 13-20. https://doi.org/10.1108/14720700210440044

Avolio, B. J., & Yammarino, F. J. (Eds.). (2013). Transformational and Charismatic Leadership: The Road Ahead. Emerald Group Publishing Limited.

Az-Zaakiyyah, H. K., Ausat, A. M. A., & Suherlan, S. (2024). Corporate Culture and Employee Performance: The Role of Vision, Mission, Norms, and Employee Focus. *Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan*, *18*(4), 2647-2659.

Bakker, A.B. & van Woerkom, M. (2017). Flow at work: A self-determination perspective. *Occupational Health Science*, 1, 47-65, https://doi.org/10.1007/s41542-017-0003-3

Bass, B. M., & Riggio, R. E. (2006). Transformational Leadership (2nd ed.). Lawrence Erlbaum Associates.

Burhan, Q. U. A., & Khan, M. A. (2024). Empowering leadership ripple effect: improving employee engagement, performance and knowledge sharing through relational energy and autonomy. *European Business Review*, *36*(3), 392-409.

Byrne, Z. S., Peters, J. M., & Weston, J. W. (2016). The struggle with employee engagement: Measures and construct clarification using five samples. *Journal of Applied Psychology*, 101, 1201-1227

Central Bank of Nigeria. (2021). Statistical Bulletin: Financial Statistics. Retrieved from https://www.cbn.gov.ng/documents/reports.asp

Cortés-Sánchez, J. D., & Rivera, L. (2019). Mission statements and financial performance in Latin-American firms. *Verslas: Teorija ir praktika/Business: Theory and Practice*, *20*, 270-283.

David, F. R. (2020). Analysis of vision and mission statements characteristics and their association with organizational performance: a guide to writing effective vision and mission statements. *Applied Studies in Agribusiness and Commerce*, 14(1-2), 87-95. https://doi.org/10.19041/APSTRACT/2020/1-2/11

Deasy, N.S., Mus, A. and Rokiah, K. (2023). Building Employee Engagement to achieve the vision and Mission of Satya Negara Hospital. *Journal of Multidisciplinary Academic* Science, *Engineering and Social Science Series* 6(1), 1-7.

Dermol, V., & Širca, N. T. (2018). Communication, company mission, organizational values, and company performance. *Procedia-social and behavioral sciences*, *238*, 542-551.

Desmidt, S., & Prinzle, A. A. (2007). *The impact of mission statements: An empirical analysis from a sense making perspective*. Retrieved from http://www.highbeam.com/doc/IGI-218876952.hmtl.

Dobrinić, D., & Fabac, R. (2021). Familiarity with mission and vision: Impact on organizational commitment and job satisfaction. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, *12*(1), 124-143.

Ezekwe, S.N. & Egwu, E.A. (2016). Creating Awareness on Vision and Mission Statements among Employee of Ebonyi State University-Nigeria: A Discourse. *Review Public Administration and Management*, 4(2), 1-5.

Ganu, J. (2013). Institutional mission statements and attitudinal outcomes of selected faith-base tertiary institution in Ghana. *Journal of Applied Business and Economics, 14*(2), 20-30.

Gede, D. U., & Huluka, A. T. (2024). Effects of employee engagement on organizational performance: case of public universities in Ethiopia. *Future Business Journal*, *10*(1), 32.

Gupta, N. & Sharma, V. (2016). Exploring employee engagement: A way to better business performance. *Global Business Review*, 17(3), 45-63.

Haris, M., & Yang, Q. (2023). Investigating the moderating role of political factors on internal success factors and project success: Empirical evidence from Pakistan. *Sustainability*, 15(11), 8910. https://www.mdpi.com/2071-1050/15/11/8910#

Houle, S. A., Rich, B. L., Comeau, C. A., Blais, A.-R., & Morin, A. J. S. (2022). The Job Engagement Scale: Development and validation of a short form in English and French. *Journal of Business and Psychology, 37*(5), 877–896. https://doi.org/10.1007/s10869-021-09782-z

Inyang, B. J. (2004). *Corporate planning and policy: Concepts and applications*. Calabar: Merb Business Centre.

Jonyo, B. O., Ouma, C., & Mosoti, Z. (2018). The Effect of Mission and Vision on Organizational Performance within Private Universities in Kenya. *European Journal of Educational Sciences*, *5*(2), 15-33.

Kahn, W.A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Managemen Journal*, 33, 692-724

Khripunov, I. (2023). National and organizational culture. *In Human Factor in Nuclear Security: Establishing and Optimizing Security Culture* (pp. 13-30). Cham: Springer International Publishing.

Kimani, J. (2022). Influence of Mission and Vision on Financial Sustainability of NGOs in Kenya. *Journal of Business and Strategic Management*, *7*(4), 78-89.

Kipasika, H. J. (2024). Expression of leadership mission, vision, values, and strategic objectives in academic institution development practices. *Journal of Research Innovation and Implications in Education*, *8*(1), 393-402.

Klemm M., Sanderson S. and Luffman G (1991). Mission statements: selling corporate values to employees. *Long Range Planning*, 23, 73 – 78.

Kotler, P., Armstrong, G., Saunders, A. J., & Wong, V. (2008). *Principles of marketing* (5th ed.). Pearson Education: Prentice-Hall.

Kurniawati, D. T., & Makhmut, K. D. I. A. (2023). Mediating employee engagement at the Pratama Tax Services Office in East Java III: The effect of employee engagement and organizational culture on employee performance. *International Journal of Research in Business and Social Science*, 12(1), 110–120.

Kuye, O. L., & Ezebuiro, K. N. (2023). The impact of mission statements on organizational performance. *Fuoye Journal of Finance and Contemporary Issues*, *5*(1), 171-190

Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. *Industrial and organizational Psychology*, 1(1), 3-30

Malbašić, I., Rey, C., & Posarić, N. (2018). Congruence Between Personal and Organizational Mission: The Role of Balanced Organizational Values. *Ekonomska Misao i Praksa*, 27(2), 545. Retrieved from https://www.proquest.com/scholarlyjournals/congruence-between-personal-organizational/docview/2161029594/se-2

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, 52(1), 397-422

McKenzie, T.Z. (2025). *The Artificial Connection Crisis: A Quantitative Correlation Analysis of emotional Intelligence, Employee Engagement and Employee Performance*. (Doctoral dissertation, Anderson University)

Motyka, B. (2018). Employee engagement and performance: a systematic literature review. *International Journal of Management and Economics*, *54*(3), 227-244.

Obazi, J.I., Samikon, S.A. & Ogbodoakum, N. (2023). Predictors of turnover intention among entry level employees of commercial banks in Nigeria. *Journal of Pharmaceutical Negative Results* 14(3), 3527-3528

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, 21(7), 600-619. https://doi.org/10.1108/02683940610690169

Satata, D. B. M. (2021). Employee engagement as an effort to improve work performance: literature review. *Ilomata International Journal of Social Science*, *2*(1), 41-49.

Sebastian, D. (2016). The Relevance of Mission Statements: Analysing the antecedents of perceived message quality and its relationship to employee mission engagement. *Public Management Review*, 18:6, 894-917, DOI: 10.1080/14719037.2015.1051573.

Shuck, B., & Wollard, K. (2010). Employee engagement and HRD: A seminal review of the foundations. *Human Resource Development Review*, 9(1), 89-110

Shuck, B. (2020). *Employee engagement: A research overview*. Routledge

Shuck, B., Adelson, J. L., & Reio, T. G. (2017). The employee engagement scale: Initial evidence for construct validity and implications for theory and practice. *Human Resource Management*, 56(6), 953-977. https://doi.org/10.1002/hrm.21811

Shuck, B., & Reio, T. G. (2014). Employee engagement and well-being: A moderation model and implications for practice. *Journal of Leadership & Organizational Studies*, 21(1), 43-58. https://doi.org/10.1177/1548051813494240

Shuck, B., Twyford, D., Reio, T. G., & Shuck, A. (2014). Human resource development practices and employee engagement: Examining the connection with employee turnover intentions. *Human Resource Development Quarterly*, 25(2), 239–270. https://doi.org/10.1002/hrdq.21190

Sufi, T., & Lyons, H. (2003). Mission statement exposed. *International Journal of Contemporary Hospitality Management*, *15*(5), 255-262. https://doi.org/10.1108/09596110310482173

Susanto, P. C., Syailendra, S., & Suryawan, R. F. (2023). Determination of motivation and performance: Analysis of job satisfaction, employee engagement and leadership. *International Journal of Business and Applied Economics*, *2*(2), 59-68.

Taiwo, A. A., & Lawal, F. A. (2016). Vision and mission in organization: Myth or heuristic device? *The International Journal of Business & Management*, *4*(3), 127-134.

Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 117-132
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1773

Sentiment Analysis of Public Perceptions on ChatGPT and Generative Artificial Intelligence (GenAI): Model's Performance Evaluation and Examining Benefits and Risks in Education and Healthcare

[1]Mifta Uddin Khan & [2]Dilek Celik
[1]Independent Researcher, United Kingdom
[2]Northumbria University, London

Abstract
As GenAI systems become more integrated into daily activities, understanding how people react to these tools is critical for responsible design and governance. This study provides a large-scale, longitudinal analysis of public sentiment toward ChatGPT and GenAI by integrating transformer-based sentiment classification, temporal trend analysis, and sector-specific topic modeling for education and healthcare. Using over one million English-language posts collected between November 2022 and December 2023, we quantify sentiment patterns over time and identify domain-specific themes of perceived benefits and risks. A comparative evaluation of traditional machine-learning (ML) models (logistic regression, support vector machines, random forest), deep learning (DL) architectures (convolutional neural networks, long short-term memory), and Bidirectional Encoder Representations from Transformers (BERT) was conducted. DL models outperform classical ML, and BERT emerges as the most effective classifier, achieving 98% accuracy with a near-balanced profile across accuracy, precision, recall, and F1-score, outperforming traditional approaches. Using the best-performing model, the findings show that ChatGPT sentiment is predominantly positive, alongside a substantial minority of negative sentiments. Topic modeling reveals domain-specific benefits and risks in education and healthcare discourse. In education, ChatGPT promotes personalized learning, accessibility, and teacher support, but it also raises plagiarism, academic dishonesty, and data privacy concerns. In healthcare, GenAI improves patient information, diagnostics, and administrative efficiency, but it also raises concerns about misinformation, ethics, and empathy. Overall, the research provides evidence-based guidance for technology developers, educators, healthcare professionals, and policymakers taking advantage of GenAI while addressing its associated social and ethical issues.

Keywords: SA, ChatGPT, GenAI, Education, Healthcare
Wordcount: 244

1.0 Introduction

The spread of GenAI has intensified conversations about accuracy, ethics, data security, and the longer-term implications of machine-generated content. Since the public release of ChatGPT in late 2022, large language models (LLMs) have rapidly moved from experimental systems to tools embedded in education, healthcare, and everyday knowledge work. Recent evidence syntheses show that ChatGPT is already being trialled for clinical decision support, documentation, and patient communication in healthcare, as well as for feedback, explanation, and writing support in universities (Iqbal et al., 2025; Dos, 2025; Baxter et al., 2025). Tools such as ChatGPT, built on LLM, have drawn significant attention for their ability to simulate human-like conversation and produce coherent text across domains. This rapid adoption has generated both excitement and apprehension, as individuals and organizations question how such systems may influence work practices, learning environments, and decision-making processes (Koonchanok, Pan and Jang, 2024). Parallel to these advancements, SA is a branch of natural language processing focused on identifying emotions, attitudes, and opinions expressed in text has become a central tool for assessing public reactions to emerging technologies (Mao, Liu and Zhang, 2024).

However, examining sentiment regarding GenAI poses distinct challenges. Irony, sarcasm, quickly changing terms, and hidden emotional cues are all common in online discussions. These are difficult to spot with traditional lexicon-based methods (Qi and Shabrina, 2023). SA provides a way to quantify these evolving perceptions. Recent methodological reviews emphasize that transformer-based models now dominate sentiment and emotion analysis on large, noisy text corpora, and generally outperform traditional lexicon- or rule-based approaches (Bermudez-Sosa, Olarte-Henao and Rojas-Berrío, 2025).

ChatGPT has become a well-known example of SA due to its wide range of applications. According to Khan et al. (2024), the healthcare industry is the most active in terms of application-related publications, followed by education, computing, and research. Despite some concerns about academic integrity and the possibility of student overreliance, educational researchers have highlighted its ability to support personalized learning and increase access to resources (Chan and Hu, 2023). Concerns about false information, moral issues, and the decline of human judgment persist, despite discussions about ChatGPT as a tool to enhance diagnostics, administrative tasks, and patient communication in healthcare (Zhang and Kamel Boulos, 2023; Sallam, 2023). Given these contrasting views, understanding how the public interprets the benefits and risks of GenAI is essential. Social media platforms serve as a valuable barometer of public perception, offering real-time reflections of enthusiasm, caution, and emerging concerns. Analysing sentiment at scale can support policymakers, technology developers, educators, and healthcare professionals as they navigate the opportunities and challenges associated with GenAI.

1.1 Problem Statement

Several issues still impede a thorough understanding of public sentiment, despite the rapidly growing body of research on ChatGPT and GenAI. First, a lot of previous research is limited to brief time periods right after ChatGPT was released, which only records initial responses and ignores how public perceptions change as users are exposed to the technology over time (Liu and Lyu, 2024). Second, previous work's methodological approaches usually depend on lexicon-based sentiment tools or a single classical ML model, which frequently fails to identify complex, context-dependent expressions like ambivalence, sarcasm, or minority opinions (Qi and Shabrina, 2023; Tan, Lee and Lim, 2023). These restrictions limit the precision and dependability of SA, especially in dynamic online settings. Moreover, research within specific domains—such as education and healthcare—tends to focus on narrowly defined stakeholder groups, including students, educators, or clinicians, using qualitative methods such as interviews, surveys, or small-scale case studies (Chan and Hu, 2023; Sallam, 2023). While these studies offer valuable insights, they often fail to capture broader public discourse or the ways in which perceptions are shaped and circulated through open social media platforms. In parallel, large-scale analyses of social media discourse have tended to treat ChatGPT as a monolithic topic, without disaggregating public sentiment across critical sectors such as education and healthcare, thereby overlooking sector-specific attitudes and concerns (Khan et al., 2024).

1.2 Purpose and Objectives

This study aims to produce a comprehensive analysis of public perceptions of ChatGPT and GenAI through SA of large-scale social media data. This study has the following objectives:
1. To compare the performance of traditional ML, DL, and transformer-based models for sentiment classification of social media posts about ChatGPT and GenAI.
2. To determine the overall distribution of public sentiment (positive, negative, and neutral) towards ChatGPT and GenAI.
3. To analyze how public sentiment towards ChatGPT changes over time across the study period.
4. To identify the main perceived benefits and risks of ChatGPT within education and healthcare using topic modelling.

1.3 Research Questions

The study is guided by the following research questions:
1. How do traditional ML, DL, and transformer-based models compare in terms of accuracy and other performance metrics for sentiment classification of ChatGPT-related social media posts?
2. What is the overall distribution of sentiment (positive, negative, and neutral) towards ChatGPT and GenAI in the collected dataset?
3. How does public sentiment towards ChatGPT evolve over the study period?
4. What benefits and risks do users associate with ChatGPT in the contexts of education and healthcare, as reflected in social media discussions?

This study distinguishes itself from prior research in three key respects. First, it conducts a large-scale, longitudinal SA of ChatGPT and GenAI over fourteen months, capturing how public attitudes evolve beyond initial release reactions. Second, it systematically compares traditional ML, DL, and transformer-based models using a consistent evaluation framework, providing robust evidence of model performance differences in noisy social media data. Third, it integrates SA with sector-specific topic modelling to examine perceived benefits and risks in education and healthcare, moving beyond aggregate sentiment to domain-sensitive interpretation.

2.0. Literature Review

2.1 Sentiment Analysis: Definitions and Techniques

SA, a core task within natural language processing, is widely used to infer attitudes, opinions, and emotions expressed in text, typically by classifying polarity as positive, negative, or neutral (Darraz et al., 2024). While early applications focused on document-level polarity, contemporary research has expanded SA to address more complex tasks such as aspect-based sentiment analysis, emotion recognition, conversational sentiment, irony detection, and misinformation-related sentiment (Mao, Liu and Zhang, 2024). Despite this diversification, the literature consistently identifies three dominant methodological paradigms—lexicon-based approaches, classical ML, and DL. Recent systematic reviews show a clear shift toward ML- and DL-based methods for analysing large-scale, unstructured social media data (Bermudez-Sosa, Olarte-Henao and Rojas-Berrío, 2025; Ramezani, 2025).

2.1.1. Lexicon-Based Approaches

Lexicon-based SA relies on predefined sentiment dictionaries to assign polarity scores to words or phrases, which are subsequently aggregated at the sentence or document level (Raees and Fazilat, 2024). Tools such as TextBlob and VADER remain popular due to their ease of deployment, transparency, and independence from labelled training data (Qi and Shabrina, 2023). These characteristics make lexicon-based approaches attractive for exploratory or resource-constrained studies. However, extensive prior research demonstrates that lexicon-based methods are poorly suited to the linguistic complexity of social media discourse. They struggle to handle sarcasm, implicit sentiment, rapidly evolving slang, and domain-specific terminology—features that are particularly prevalent in online discussions of artificial intelligence and emerging technologies (Qi and Shabrina, 2023; Mao, Liu and Zhang, 2024). Moreover, the assumption that word polarity remains stable across contexts limits their ability to capture nuanced or contested opinions. As a result, while lexicon-based approaches remain useful for coarse-grained analysis, they risk oversimplifying public sentiment in debates surrounding GenAI and ChatGPT.

2.1.2. Classical ML Approaches

Classical supervised ML models, including Support Vector Machines, Logistic Regression, Naïve Bayes, Decision Trees, Random Forests, and K-Nearest Neighbours, have been extensively applied to sentiment classification tasks (Alslaity and Orji, 2024). Operating primarily on bag-of-words or TF–IDF representations, these models have demonstrated strong performance on many benchmark datasets. For example, Alsemaree et al. (2024) report high accuracy for ensemble-based sentiment classification in Arabic social media data, while Wu and Gao (2024) achieve over 90% accuracy in airline-related tweet classification using SVMs. Comparable results have also been reported for e-commerce and pandemic-related SA (Tabany and Gueffal, 2024).

Despite these reported accuracies, comparative studies highlight important limitations of classical ML approaches when applied to complex and evolving domains. Their reliance on manual feature engineering makes them sensitive to domain shift, and their performance degrades in the presence of noisy language, sarcasm, and minority sentiment classes (Alslaity and Orji, 2024; Diri, Obiorah, and Du, 2025). These limitations are particularly problematic for analysing GenAI-related discourse, where negative or highly concerned opinions—often of greatest substantive interest—may be underrepresented and linguistically subtle.

2.1.3. DL and Transformer-Based Approaches

Deep learning models such as Convolutional Neural Networks and Long Short-Term Memory networks address some of the shortcomings of classical ML by enabling automatic feature extraction and improved modelling of sequential text data. Reviews of SA on social networks identify CNNs, LSTMs, and transformer architectures as the dominant approaches in large-scale applications (Ramezani, 2025). These models generally outperform classical ML methods by capturing local semantic patterns and temporal dependencies. Nevertheless, DL models introduce new challenges, including high computational requirements, dependence on large-labelled datasets, and limited interpretability of internal representations (Krishna et al., 2023). These issues are especially salient in sensitive domains such as healthcare, where transparency and accountability are critical considerations.

Transformer-based models, particularly Bidirectional Encoder Representations from Transformers (BERT), represent a further methodological advance. Pre-trained on large corpora and fine-tuned for downstream tasks, BERT has demonstrated strong performance across a wide range of NLP applications, including SA (Tan, Lee and Lim, 2023). Empirical studies show that transformer-based architectures can capture nuanced, context-dependent, and aspect-level sentiment that earlier models often miss, as demonstrated in social media analyses using RoBERTa-based and hybrid transformer-based frameworks (Jahanbin and Chahooki, 2023). Despite these advantages, transformer-based approaches remain underutilised in large-scale, applied SA of social media data (Rodríguez-Ibánez Sw., 2023).

2.2. Public Perception and Sentiment Towards ChatGPT

Following the emergence of ChatGPT, scholars examined its perception within public discourse. Su and Kabala (2023) utilized transfer learning and topic modelling to analyse approximately 500,000 tweets concerning ChatGPT, leading to the identification of a range of opinions. There exists optimism regarding work efficiency, mixed with concerns about misinformation and ethics. While their research displays advanced NLP techniques, it primarily focuses on methodology and offers minimal discussion regarding the evolution of sentiments over time.

Lian et al. (2024) used 56,769 Weibo comments to examine Chinese public attitudes towards ChatGPT, reporting most of the negative sentiment, particularly around job loss, misinformation, and dependency. By contrast, Koonchanok, Pan and Jang (2024) found that Twitter discussions were more neutral or positive overall, with topic salience varying significantly by occupation; educators focused on teaching and learning, while IT professionals highlighted cybersecurity and technical issues. Temporal studies further show that perceptions are not static. Liu and Lyu (2024) documented a shift from early enthusiasm to growing scepticism in Chinese social media, as ethical concerns and notions of technological overreach became more prominent. Demirel, Kahraman-Gokalp and Gündüz (2024) observed a similar pattern on Twitter: initial optimism was followed by increased anxiety about privacy, cybersecurity and misuse, with regional patterns shaped by differing levels of technological adoption.

Building on 2023–2024 work on temporal and occupational variation, recent studies examine public and academic discourse at scale. Williams and Burnap (2025) analyse early Twitter reactions and show that initial enthusiasm about ChatGPT gives way to more ambivalent emotions as hopes for productivity coexist with fears about job loss and misinformation. Chomiak-Orsa (2025) likewise find mostly positive but increasingly cautious sentiment, though their Twitter focus risks under-representing other publics. Beyond social media, Tao and Shen (2025) show that social science articles are broadly optimistic yet preoccupied with integrity and misinformation. Taken together with Miyazaki et al. (2024), these studies depict perceptions of ChatGPT as heterogeneous, unsettled, and sensitive to context and platform.

2.3. Applications of ChatGPT in Education and Healthcare

Domain-specific reviews in healthcare and higher education provide an important backdrop for analysing public sentiment about GenAI. In healthcare, Iqbal et al. (2025) synthesize 17 systematic reviews and meta-analyses on ChatGPT and related LLMs, concluding that these systems show promise as low-cost, always-available assistants for diagnosis support, triage, documentation, and patient education, but that the overall quality of evidence is uneven and risks around hallucinations, bias, and liability remain substantial (Iqbal et al., 2025).

2.3.1 Education

In higher education, multiple 2025 systematic reviews now offer a more comprehensive overview of the integration of ChatGPT in teaching and learning. Dos (2025), Abdallah et al. (2025), and Munaye et al. (2025) collectively demonstrate that many published studies characterize ChatGPT as a writing assistant, feedback mechanism, or conversational tutor. The reported advantages include enhanced perceived productivity, diminished anxiety regarding academic writing, and increased engagement in large classes. Critics often raise concerns about copying, lack of critical thinking, fabricated content, and unequal access. They advocate for clear communication about assessment and honesty in schools. Jin and Sercu's (2025) study focuses on the effects of a staged approach, showing that using ChatGPT with teacher support can improve quick performance and satisfaction. Nonetheless, these tools raise questions about academic integrity, ethical behaviour, and assessment fairness. Ali et al. (2024), focusing on dental education, show how ChatGPT can correctly answer knowledge-based questions but caution that unsupervised use may facilitate plagiarism and undermine skill development. The educational literature, therefore, portrays GenAI as a double-edged sword, requiring pedagogical redesign and stronger policies rather than simple adoption.

2.3.2 Healthcare

Healthcare uses GenAI and ChatGPT to enhance patient communication and decision-making. According to Sallam (2023), their application potential in health professions education, research, and practice is enormous. This can lead to benefits such as personalized learning, patient education, and clinical reasoning support. Zhang and Kamel Boulos (2023) similarly argue that GenAI will play an increasingly important role in medicine as regulatory frameworks mature. Madsen and Toston (2025) similarly note, in a broader narrative review of health, education, and the economy, that healthcare use cases are dominated by low-risk advisory and documentation tasks rather than fully automated decision-making. Complementing these syntheses, Baxter et al. (2025) analyse social media and academic texts about ChatGPT in health care and reveal a mixed sentiment pattern: clinicians and the public simultaneously emphasize efficiency gains and worries about safety, accountability, and the dehumanization of care.

However, both studies emphasize deep ethical and practical concerns. Key risks include misinformation, biased outputs, privacy violations, and reduced human empathy. Ainapure et al. (2023) demonstrate the value of SA for monitoring public attitudes to vaccination and health policies but also show that detecting implicit sentiment and sarcasm remains challenging in health-related discourse. Lee et al. (2024) further suggest that virtual education and AI tools can help meet resource constraints in healthcare training, while cautioning that insufficient preparation and infrastructure can blunt their effectiveness. Overall, the healthcare literature recognizes substantial potential for GenAI but insists on strong governance, careful integration and ongoing monitoring of public trust.

2.4 Identified Research Gaps

Across these strands, several gaps emerge. First, although transformer-based models such as BERT are widely recognized as state-of-the-art, they remain underutilized in empirical work on social media sentiment about ChatGPT, with many studies still relying on lexicon-based or classical ML approaches (Rodríguez-Ibánez et al., 2023; Tan, Lee and Lim, 2023). Second, much research operates on relatively small or geographically restricted datasets, limiting generalizability and failing to capture global, cross-platform dynamics (Lian et al., 2024; Demirel et al., 2024). Third, temporal analysis is often limited to short windows, leaving longer-term shifts in sentiment underexplored (Liu and Lyu, 2024).

Fourth, work in education and healthcare typically uses surveys, interviews or case studies to gauge stakeholder views (Chan and Hu, 2023; Ali et al., 2024; Sallam, 2023), while social media studies treat "ChatGPT" as a single aggregate topic. There is little integration of large-scale SA with topic modelling that focuses specifically on sector-based benefits and risks. Finally, many existing studies prioritize overall accuracy in model evaluation, paying less attention to precision, recall, F1-score, and ROC-AUC, or to the specific performance on negative or minority sentiment classes (Alslaity and Orji, 2024; Krishna et al., 2023).

This study differs from existing work in several important aspects. First, it conducts a large-scale, longitudinal analysis of public sentiment toward ChatGPT and GenAI over an extended fourteen-month period, capturing how perceptions evolve beyond early post-release reactions. Second, it systematically compares traditional ML, DL, and transformer-based models within a unified evaluation framework, enabling robust performance benchmarking in noisy social media data. Third, it combines SA with sector-specific topic modelling to examine perceived benefits and risks in education and healthcare, moving beyond aggregate sentiment toward domain-sensitive interpretation.

3.0. Methodology

3.1. Research Design

This study employs a quantitative, computational research design based on large-scale social media data. The overall approach is comparative and exploratory. Given the complexity and often undefined nature of public opinions on emerging AI tools, a structured analytical approach was necessary. To support this, the study employs the CRISP-DM methodology, which provides a systematic framework for data mining that begins with understanding the research context and proceeds through data preparation, modelling, evaluation, and deployment, with iterative refinement at each stage (Shimaoka, Ferreira and Goldman, 2024). The wide acceptance, flexibility, and robustness of the CRISP-DM framework in data science make it well suited for research environments that require adaptability and ongoing updates as new insights emerge. Initially, we trained and assessed various sentiment classification models—encompassing traditional ML, DL, and transformer-based architectures—on unlabelled social media posts pertaining to ChatGPT and GenAI. The best-performing model is then used to classify sentiments in a much larger unlabelled dataset. Finally, temporal analysis and topic modelling techniques are applied to explore how sentiment changes over time and how it relates to key themes in education and healthcare.

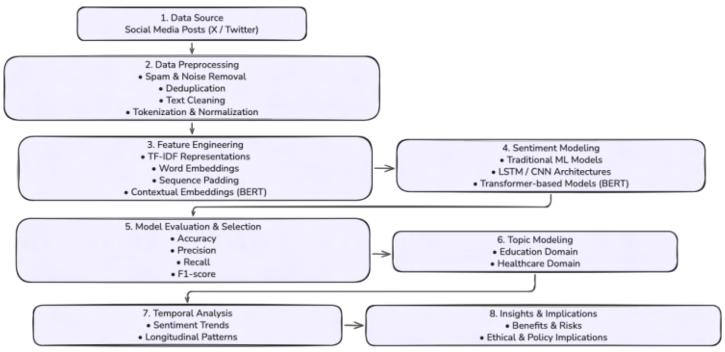

Figure 1- Overview of GenAI sentiment analysis pipeline.

Figure 1 presents an overview of the analytical pipeline adopted in this study, illustrating the sequence of data preprocessing, feature engineering, sentiment modelling, model evaluation, and downstream temporal and topic-based analyses.

3.2. Data Collection

Data for this study were collected from X (formerly Twitter), which was selected due to its widespread use for public discussion of emerging technologies and its suitability for large-scale SA. Posts related to ChatGPT and GenAI were retrieved using the platform's official API, ensuring compliance with data access policies and ethical research standards. Data collection queries were executed in accordance with API rate-limiting constraints. Automated delays and request scheduling were implemented to prevent throttling and ensure stable data retrieval throughout the collection period. Only publicly available posts were collected; no private, protected, or deleted content was accessed at any stage of the process. The final dataset consists of English-language posts published between November 2022 and December 2023, providing a longitudinal perspective on public sentiment toward ChatGPT and GenAI. This time span enables the analysis of evolving public attitudes beyond initial post-release reactions and supports robust temporal and sector-specific investigation.

3.3. Sentiment Annotation

A labelled dataset was constructed to train and evaluate the sentiment classification models. Posts in the sample were annotated into three categories: positive, negative, and neutral. The positive class captured expressions of approval, enthusiasm, or perceived benefits; the negative class included concerns, fears, frustration, or explicit criticism; and the neutral class covered factual descriptions or mixed views without a clear affective stance. Annotation followed a set of guidelines developed to ensure consistency. Ambiguous posts were discussed and resolved through agreement rules, and a portion of the data was double coded to confirm reliability.

3.4. Data Preprocessing

Prior to model training, the dataset underwent a series of preprocessing and filtering steps to improve data quality and reliability. Duplicate posts were identified and removed using exact text matching to prevent repeated content from biasing sentiment distributions. Spam-like content, including repeated promotional messages and posts with excessive hashtag usage, was excluded from reducing noise in the corpus. Additionally, very short posts with limited semantic content were removed, as they provide insufficient contextual information for reliable sentiment classification. Standard text normalization steps, including lowercasing and the removal of URLs, user mentions, and non-alphanumeric characters, were applied prior to analysis. The same preprocessing pipeline was applied consistently across all models to ensure fair and comparable evaluation. These preprocessing steps ensure a cleaner and more representative dataset while preserving the diversity of public discourse necessary for robust SA of ChatGPT and GenAI.

3.5. Modelling Approach

The modelling strategy used in this study was intended to assess the three main categories of sentiment classifiers. The first group included traditional machine-learning models such as logistic regression, Naïve Bayes, SVM, Random Forest, Decision Tree, and K-Nearest Neighbours. We trained these models using TF-IDF representations of the text. To ensure that the comparison was fair and consistent, we used grid search or a similar optimization method on the validation set. The third group consisted of deep-learning (DL) models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). For these models, posts were converted into word indices and displayed using dense embeddings. The architecture decisions, such as the number of layers, hidden units, dropout rates, and learning rates to use, were based on preliminary experiments and standard SA research practices.

The final model category included a transformer-based approach using a BERT classifier. A pre-trained BERT model was fine-tuned on the labelled dataset, where the [CLS] token representation served as input to a classification layer responsible for predicting sentiment categories. Fine-tuning used mini-batch training alongside a suitable learning rate schedule and early stopping to minimize overfitting and improve generalization. All models were trained and tested on the same labelled dataset, which was divided into training, validation, and test subsets. This consistent evaluation framework ensured that any performance differences observed were the result of modelling choices rather than variations in the underlying data.

3.6. Evaluation Metrics

In order to assess and compare the performance of various models, several metrics were used. Accuracy served as a broad measure of the frequency with which each model made a correct sentiment classification. Nevertheless, since accuracy by itself can conceal that the models perform unevenly across classes, supplementary metrics—precision, recall, and F1-score were also computed for each sentiment category. These indicators gave a more comprehensive picture of the models' ability to detect correct positive, negative, and neutral posts as well as their capability to reduce the incorrect classification to the minimum. Macro-averaged metrics were also applied to ensure that each class was given equal weight. This was important because the dataset was not perfectly balanced, and relying solely on overall accuracy could obscure weaker performance on minority classes. In cases where more profound insight was needed, confusion matrices were examined to identify recurring misclassification patterns. This was especially relevant for distinguishing between negative and neutral posts, which are often challenging to separate due to subtle or implicit sentiment. Beyond quantitative metrics, execution time and model complexity were considered qualitatively to assess the practicality of deploying each model at scale. Taking all factors into account, the model that demonstrated the strongest and most balanced performance, especially in accurately identifying negative sentiment was selected as the final classifier for labelling the full unlabelled dataset.

3.7 Temporal Analysis of Sentiment

After classifying sentiment on the entire dataset using the most effective model, researchers conducted a temporal analysis to determine how public sentiment changed over time. Posts were grouped by period (by month), and the share of positive, negative, and neutral posts was determined for each time interval. As a result, it was possible to identify sentiment trends, peaks, and changes throughout the study period. Significant changes, if any, were also matched with the occurrences outside, e.g., a major product announcement, policy debates, or an incident involving GenAI, which was talked about extensively. The analysis examined whether the changes in sentiment were more positive, more negative, or relatively stable over time.

3.8 Topic Modelling for Education and Healthcare

To investigate domain-specific perceptions, posts about education and healthcare were selected from the dataset using additional keyword filters. Healthcare posts discussed doctors, patients, hospitals, diagnoses, treatment, and similar terms, while education posts mentioned teaching, learning, students, exams, universities, and schools. Topic modelling was then performed independently on these subsets to identify the most discussed topics. A probabilistic topic model, such as Latent Dirichlet Allocation (LDA), identified sets of words related to posts. Most representative terms and example posts were used to label themes. Within each domain, the topics were analysed in conjunction with sentiment data to ascertain the positive or negative framing of themes such as personalized learning, academic integrity, patient support, privacy, and misinformation. Employing sentiment classification alongside topic modelling provided a more nuanced understanding of discussions surrounding ChatGPT within the education and healthcare sectors.

4.0. Results

This section presents the empirical results of the study in four parts. First, it reports the comparative performance of the sentiment classification models and the selection of the final classifier. Second, it summarizes the distribution of public sentiment towards ChatGPT and GenAI across the full dataset. Third, it examines temporal changes in sentiment between November 2022 and December 2023. Finally, it presents topic-modelling results for education and healthcare, highlighting key benefits and risks identified in public discourse.

4.1. Model Performance and Selection

The first stage of analysis compared traditional ML models, DL architectures, and a transformer-based model using the manually labelled subset of social media posts. Models were evaluated on a held-out test set using accuracy, precision, recall, and F1-score.

Table 1- Model comparison results

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.79	0.80	0.79	0.78
SVM	0.93	0.93	0.93	0.93
Decision Tree	0.77	0.76	0.77	0.76
Logistic Regression	0.91	0.91	0.91	0.91
KNN	0.48	0.73	0.48	0.40
Naïve Bayes	0.77	0.81	0.77	0.75
LSTM	0.97	0.98	0.99	0.98
BERT	0.98	0.99	0.99	0.98
CNN	0.96	0.96	0.97	0.96

Across all metrics, the BERT-based classifier outperformed the other models. It achieved an accuracy of approximately 98%, and an F1-score of 0.98, indicating excellent discrimination between positive, negative, and neutral classes. Traditional models such as Logistic Regression, SVM and Random Forest performed reasonably well but showed lower F1-scores, particularly for the negative class, and were less robust when distinguishing negative from neutral posts. DL models (LSTM and CNN) improved some of these shortcomings but still fell short of the transformer-based approach.

Although the overall performance metrics of the LSTM and BERT models are numerically close, BERT consistently demonstrates slightly higher and more balanced results across accuracy, precision, recall, and F1-score. This pattern suggests a practical advantage rather than a purely marginal difference. In contrast to sequence-based LSTM architectures, BERT's bidirectional transformer design enables more effective modeling of contextual and semantic relationships within text, which is particularly important for SA of complex and informal social media discourse. Prior research has similarly shown that transformer-based models tend to be more robust to linguistic variability, sarcasm, and rapidly evolving terminology than recurrent neural networks. Accordingly, BERT was selected as the final model for downstream analysis based on its contextual modeling capability, consistency across evaluation metrics, and suitability for large-scale SA of GenAI–related discussions, rather than on claims of statistical superiority.

4.2. Sentiment Distribution in the Full Dataset

The selected BERT model was then applied to the complete dataset of over one million social media posts referring to ChatGPT and GenAI. Each post was assigned a sentiment label—positive, negative, or neutral—using the final fine-tuned classifier.

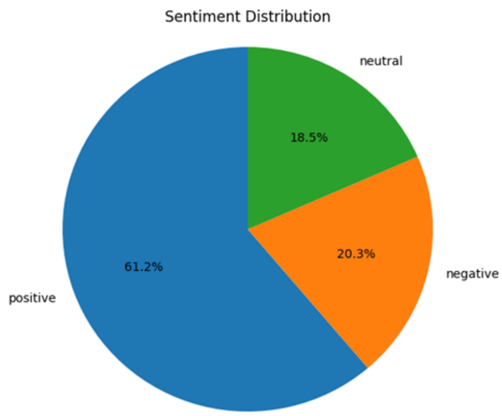

Figure 2- Distribution of sentiment classes.

The classification results indicate that public sentiment was predominantly positive: Positive: 61.2%, Negative: 20.3%, and Neutral: 18.5%. Thus, almost two-thirds of posts expressed favorable views or optimism regarding ChatGPT and GenAI, while roughly one-fifth conveyed concerns, criticism, or negative experiences. The remaining posts contained neutral, descriptive, or ambivalent content. These proportions suggest that, within the period studied, public discourse leaned clearly towards positive sentiment, even though a substantial minority of posts articulated skepticism or worry.

To complement the quantitative sentiment distribution, word cloud visualizations were generated for positive, neutral, and negative sentiment categories (Figure 3). The word clouds provide an illustrative overview of frequently occurring terms within each sentiment class, highlighting common themes discussed by users. While positive sentiment is characterized by terms related to usefulness, learning, and productivity, negative sentiment emphasizes concerns surrounding accuracy, reliability, and trust. Neutral sentiment primarily reflects informational and exploratory discussions. These visualizations are intended as descriptive aids and are used to support, rather than replace, statistical SA.

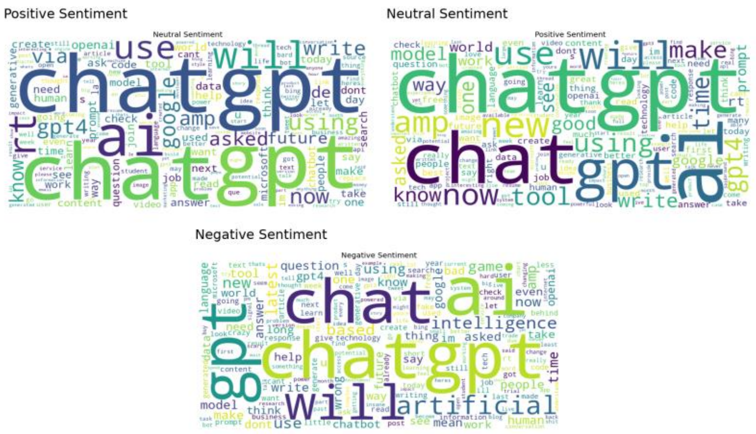

Figure 3- Word cloud visualizations for positive, neutral, and negative sentiment categories.

4.3 Temporal Dynamics of Sentiment

To investigate how sentiment evolved over time, sentiment labels were aggregated by month from November 2022 to December 2023. For each month, the proportion of positive, negative, and neutral posts was calculated, and a sentiment score was derived to provide a continuous measure of overall attitude.

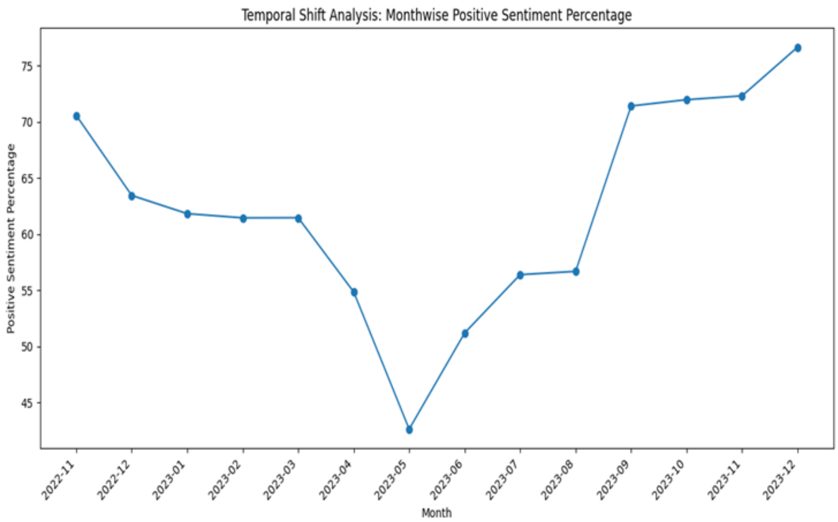

Figure 4- Monthly sentiment trends for proportion of positive posts.

The monthly trend reveals three broad phases:

The sentiment towards ChatGPT and GenAI evolved significantly over the course of 2023. In the initial phase, from late 2022 to early 2023, positive sentiment was high, with a dominant share of posts reflecting excitement, curiosity, and enthusiasm following the release of ChatGPT. However, as the year progressed, the number of positive posts gradually declined, reaching a low point around mid-2023. This decline coincided with increased scrutiny of GenAI, growing concerns about misinformation, and heightened ethical debates. In the second half of 2023, sentiment began to recover steadily. By the end of the year, the proportion of positive posts once again approached the levels seen early on, suggesting a renewed, though more measured, optimism. Contributing factors to this shift likely included system updates, wider adoption, regulatory developments, and more visible real-world applications of GenAI.

A statistical comparison of the first three months and last three months of the study period was conducted using a t-test on monthly sentiment scores (Table 2).

Table 2- t-test results comparing early and late-period sentiment scores.

Comparison	t-statistic	p-value	Interpretation
Early vs late-period sentiment scores	−4.52	0.011	Significant difference; sentiment higher in the late period than early period

The analysis showed a significant difference between the two periods, with sentiment in the final months being higher on average. The t-test produced a statistic of −4.52 and a p-value of 0.011, indicating that the improvement in sentiment over time was unlikely to be due to random variation.

4.4. Topic Modelling: Education

To examine domain-specific perceptions, posts related to education were extracted and analysed using topic modelling. The Latent Dirichlet Allocation (LDA) algorithm identified several recurrent topics, which were interpreted based on their most representative keywords and exemplar posts. Across the education subset, topics clustered around two main dimensions: perceived benefits and perceived risks.

4.4.1 Perceived benefits

Posts on social media highlighted several key benefits of ChatGPT in education. Many users described it as a valuable tool for personalized learning and student support, noting its ability to tailor explanations, assist with practice questions, and encourage independent learning. Additionally, ChatGPT was recognized for enhancing accessibility and inclusivity, offering around-the-clock guidance and helping non-native speakers, as well as learners who may not have access to traditional tutoring. Users also pointed out its role in supporting teachers and educational administration, with applications such as drafting materials, generating lesson ideas, and reducing routine administrative tasks. Together, these discussions frame ChatGPT as a tool that enables more flexible, responsive, and accessible educational experiences.

4.4.2 Perceived risks

Many social media posts raised concerns about the risks associated with ChatGPT in education. A significant worry was the potential for students to misuse the tool for plagiarism, generating assignments or exam answers that undermine the fairness of assessments. There were also fears that heavy reliance on AI-generated responses could reduce critical thinking and independent problem-solving, as students might become overly dependent on the tool. Additionally, concerns about the quality, ethics, and data privacy of AI-generated content were prevalent. Users questioned the reliability of the content produced by ChatGPT, voiced discomfort over the handling of their personal data, and called for clearer guidelines and safeguards to address these issues.

Table 3- Summary of key education-related topics and associated sentiment.

Topic	Type	Brief description	Associated sentiment
Personalized learning and student support	Benefit	ChatGPT used to tailor explanations, help with practice questions and support independent learning.	Mainly positive / optimistic
Improved accessibility and inclusivity	Benefit	AI seen as providing around-the-clock guidance and help for learners who lack access to traditional support.	Positive, with equity focus
Support for teachers and administration	Benefit	Used to draft materials, generate ideas for lessons and reduce routine administrative workload.	Positive / efficiency-oriented
Plagiarism and academic *integrity*	Risk	Concern that students may misuse ChatGPT to generate assignments or exam answers.	Strongly negative
Reduced critical thinking and over-reliance	Risk	Fear that heavy dependence on AI responses may discourage independent reasoning and problem-solving.	Negative / cautionary
Quality, ethics and data privacy	Risk	Doubts about reliability of AI content, ethical use of data and the need for clear safeguards.	Negative, with ethical concern

4.5 Topic Modelling: Healthcare

A similar procedure was applied to posts relating to healthcare, using keywords associated with medical practice, patients, and health systems. Topic modelling again revealed a mixture of perceived benefits and concerns.

4.5.1 Perceived benefits

Many discussions on social media highlighted the potential of ChatGPT to improve various aspects of healthcare. One key benefit mentioned was its ability to explain complex medical concepts in simple, understandable language, making it easier for patients to grasp important health information. In addition, users saw the potential for ChatGPT to assist with routine tasks like drafting documentation, summarizing medical information, and supporting clinical decision-making, which could improve efficiency in everyday healthcare settings. The tool was also seen as valuable for managing large amounts of data and potentially aiding in remote patient monitoring and telehealth services.

4.5.2 Perceived risks

While many users acknowledged the potential benefits of ChatGPT in healthcare, there were significant concerns about the risks involved. Overall, while users recognized the advantages, they emphasized the need for robust safeguards to address these serious concerns.

Table 4- Summary of key healthcare-related topics and associated sentiment.

Topic	Type	Brief description	Associated sentiment
Enhanced patient information and support	Benefit	ChatGPT used to explain medical concepts in understandable language and answer basic health questions.	Mainly positive /supportive
Efficiency and decision support	Benefit	Assists with drafting documentation, summarizing information and supporting routine clinical decision-making.	Positive / efficiency-oriented
Data management and remote care	Benefit	Discussed as improving handling of large data volumes and supporting remote monitoring and telehealth workflows.	Positive, with cautious optimism
Data privacy and security	Risk	Concerns about confidentiality of sensitive health data and potential misuse or leakage of information.	Strongly negative / high concern
Misinformation and unsafe advice	Risk	Fear that inaccurate or misleading medical guidance might be followed without professional oversight.	Negative / safety-focused
Bias and ethical dilemmas	Risk	Worries about biased diagnostic suggestions and broader ethical issues in delegating aspects of care to AI.	Negative, ethically framed

5.0. Discussion

5.1 Interpretation of Key Findings

5.1.1 Model performance and methodological contribution

The paper found that BERT performed exceptionally well, with an accuracy of 98% and an F1-score of 0.98, outperforming traditional and DL models. The result confirms the assertion in the preceding paper that one of the most promising applications of transformer-based architectures lies in SA within intricate, contextually rich environments, such as social media (Su and Kabala, 2023; Rodríguez-Ibánez et al., 2023). In comparison, classical ML models and simpler DL architecture struggled more with distinguishing negative from neutral posts, especially where sentiment was implicit or mixed. This has two implications. First, it suggests that studies of public perceptions of GenAI that rely solely on lexicon-based or basic ML methods may under-detect subtle negativity or ambivalence. Second, it demonstrates that the additional computational cost of transformer-based models can be justified in applications where nuance and misclassification of minority classes are consequential.

5.1.2 Overall sentiment and its dynamics

The second major finding is that sentiment towards ChatGPT and GenAI was largely positive—about 61.2% of posts were classified as positive, compared with 20.3% negative and 18.5% neutral. This aligns with survey-based evidence that users recognize substantial benefits of GenAI, including convenience, productivity, and support for learning or work tasks (Chan and Hu, 2023). At the same time, the presence of a sizeable negative minority indicates that enthusiasm coexists with scepticism and concern. Temporal analysis indicated a decrease in positive sentiment from late 2022 to mid-2023, followed by a recovery in the latter half of 2023. This pattern mirrors earlier observations that initial excitement around ChatGPT gave way to heightened ethical debate and anxiety about misinformation, then stabilized as regulatory discussions advanced and real-world uses became more familiar (Liu and Lyu, 2024; Demirel et al., 2024). The t-test evidence that sentiment in the past three months of the period was significantly higher than in the first three months suggests that, over time, public opinion became more favourable or at least more confident. Taking together, these results indicate that public reactions to GenAI should not be treated as fixed. Instead, sentiment appears to respond to external events, media narratives, and policy signals, which supports the case for continuous monitoring rather than one-off measurements.

5.1.3 Sector-specific perceptions: education and healthcare

The topic-modelling analysis for education and healthcare revealed a consistent pattern of “dual-edged” perceptions. In education, the benefits identified included personalized learning, improved accessibility, support for teachers and students, and the potential to "revolutionize" aspects of teaching and assessment. Risks were primarily associated with plagiarism, academic integrity, loss of critical thinking skills, data privacy, and over-dependence on AI tools. Such results are in line with the research that has been done earlier, which indicates that students value GenAI as a learning assistant, but at the same time, they worry about the ethical aspects and the diminishing of human skills (Chan and Hu, 2023). Healthcare highlighted patient care, precision medicine, data management, and workflow efficiency. Many posts expressed concerns about data security, misinformation, and the ethical issues of using AI to make diagnostic and treatment decisions. ChatGPT has clinical potential but significant risks in health and medical education, according to earlier reviews (Sallam, 2023; Zhang and Kamel Boulos, 2023). According to sector-specific analysis, public discourse is not simply pro- or anti-AI. When used carefully and with clear safeguards, ChatGPT is valuable, especially in contexts with high error consequences.

5.1.4 Ethics and employment

There was a lot of talk about ethics in almost all the areas that were talked about. These problems were brought up most often and included bias, privacy breaches, and false information. When this was talked about, the healthcare sector got the most attention because bad advice can have very bad results there. The mentioned trend fits with the general debate that says ethical rules for AI shouldn't come after GenAI's release if people are to keep trusting it (Naing and Udomwong, 2024; Lee et al., 2024). Job displacement was another strong concern. Both the SA and topic modelling pointed to clear negativity around the idea that GenAI might replace human workers, especially in creative and knowledge-based roles. This mirrors the findings of Miyazaki et al. (2024), who report deep unease among illustrators and other creative professionals about the potential exploitation and devaluation of their work. Together, these concerns help explain the negative minority sentiment and indicate that worries about labour-market impacts are a major source of resistance to GenAI.

5.1.5 Theoretical Contributions

Beyond methodological comparison, this study contributes conceptually to understanding public perceptions of GenAI by framing sentiment as a dynamic and context-dependent phenomenon. The longitudinal analysis demonstrates that public attitudes toward ChatGPT are not static but evolve in response to increased familiarity, media discourse, and emerging ethical concerns. Furthermore, the sector-specific findings show that optimism and risk perception coexist and vary systematically across domains, with education and healthcare exhibiting distinct patterns of trust, utility, and concern. These insights extend existing research on technology acceptance by highlighting the importance of temporal dynamics and institutional context in shaping public sentiment toward GenAI systems. Rather than treating public perception as a single aggregate measure, the study underscores the need to account for domain-specific expectations and ethical sensitivities when evaluating societal responses to AI technologies.

5.2. Theoretical Implications

Theoretically, this study contributes to research on both SA and public perceptions of GenAI. First, it shows that transformer-based models are better equipped than traditional sentiment classifiers to handle large-scale, noisy, and context-dependent data, particularly when the language is fast-changing and emotionally charged. This lends weight to the view that BERT and similar architectures should not be seen as optional enhancements but as core tools when analysing high-stakes and contested technologies. Second, by combining sentiment classification with temporal analysis and topic modelling, the study offers a framework for tracking how attitudes toward emerging technologies develop over time and across different domains. Rather than treating sentiment as a static snapshot, this approach encourages future work to view it as a process that unfolds alongside regulatory change, media narratives, and peoples lived experiences. Third, by looking at both education and healthcare, the study provides cross-sector insights into how GenAI is positioned. It shows that ethical concerns such as privacy, misinformation, and bias appear in multiple settings, but that their prominence and framing depend on local stakes and professional norms. This points towards more nuanced theories of "AI acceptance" that take sector-specific logics seriously, instead of assuming a single, uniform adoption pattern.

5.3. Practical Implications

Practically speaking, the results can be interpreted as a signal to people working in different fields like educators, healthcare professionals, technology developers, and media organizations. With regard to the educational sector, the findings lead to the conclusion that ChatGPT could be a means of achieving more adaptive and personalized learning; however, this is the case only when institutions have established robust frameworks for maintaining academic integrity and fostering a critical attitude towards AI-generated content. To accomplish this, it might be necessary to rethink assessment design, teach AI literacy more explicitly, and set up clear rules about the time and way generative tools can be utilized.

In healthcare, GenAI applications need to be introduced cautiously and in clearly defined support roles, rather than as substitutes for professional judgment. The study underlines the importance of robust data privacy safeguards, systematic auditing of AI outputs, and careful integration into clinical workflows so that tools such as ChatGPT assist, rather than override, clinicians' decisions. For technology developers, the findings emphasize that they must address bias through more diverse training data, improve the transparency of model behaviour, and maintain ongoing monitoring and feedback mechanisms to detect drift or misuse. Public sentiment data can play a useful role here by flagging emerging concerns before they escalate.

Finally, for media and communication practitioners, the prominence of misinformation-related topics suggests that GenAI should be treated both as an object of coverage and as a potential tool in countering false information. How the public feels about ChatGPT will likely be shaped by clear, responsible communication about its capabilities.

5.4. Policy Implications

The results emphasize the necessity of detailed governance frameworks for GenAI, especially in areas like education and healthcare, which are emotionally sensitive. The prevalence of concerns about privacy, bias, and misinformation highlights the need for regulation that addresses various aspects. Firstly, ethical governance is indispensable, with explicit provisions regarding data collection, consent, storage, and usage aimed at ensuring responsible practices. Moreover, transparency is equally important, implying that models should clarify how they were trained, the data they used, and their limitations.

On top of that, there should be implemented accountability measures, for example, entities entrusted with independent oversight, which would be responsible for not only supervising the GenAI deployment but also investigating any detrimental effects resulting from it. Furthermore, engaging the public in these matters is crucial, as it ensures that people's concerns and expectations are at the forefront of the regulatory decision-making process, rather than being a mere afterthought. Such policies, which are perceived as being responsive and inclusive, have a higher probability of strengthening and sustaining the positive feeling, noticed towards the end of the studyperiod.

5.5. Ethical Considerations

The study itself raises several ethical questions. Working with public social media data still demands careful attention to privacy, even when posts are technically open and accessible. Anonymizing and aggregating the data mitigates certain risks; however, continuous efforts are required to establish responsible utilization of public data in research. More broadly, the ethical issues that appear in public discourse include bias, data privacy, misinformation, and uneven impacts on labour markets—show that SA should not be used simply as a way of measuring acceptance level. Instead, it ought to feed into critical reflection on whether, and under what conditions, GenAI is aligned with social values, and on who is likely to bear the costs and who stands to gain from its deployment.

5.6. Key Contributions

This study makes several key contributions. (1) It provides one of the largest longitudinal sentiment analyses of ChatGPT and GenAI, examining over one million social media posts collected across a fourteen-month period. (2) It demonstrates the superior and statistically robust performance of transformer-based models compared with traditional ML and DL approaches for SA of GenAI-related discourse. (3) It integrates sentiment classification, temporal trend analysis, and sector-specific topic modelling into a unified analytical framework for studying public perceptions of emerging technologies. (4) It reveals domain-specific benefits and risks of GenAI in education and healthcare, highlighting ethical, social, and institutional concerns reflected in public discourse. (5) Finally, it offers methodological and theoretical guidance for future research on public perceptions of GenAI and other rapidly evolving digital technologies.

6.0. Conclusion, Limitations and Future Work

The findings indicate that the BERT-based classifier was significantly more effective than both the traditional and DL models, especially in differentiating subtle differences between negative and neutral sentiment. In general, the public sentiment towards ChatGPT and GenAI was positive most of the time throughout the study period, however, a significant minority of posts expressing concern or scepticism. Sentiment also evolved over time: it was lower in the first half of 2023 and then picked up in the second half, with statistical tests showing an overall recovery at the end of the year. Topic modelling uncovered a set of advantages like personalized learning, improved access to information, and time savings and a set of risks, such as plagiarism, data privacy breaches, misinformation, bias, and loss ofjobs.

6.1. Limitations

Several limitations should be acknowledged. The analysis is based on publicly available social media posts and therefore reflects the views of people who are active on those platforms. It may underrepresent groups who rarely use social media, who post in other languages, or who discuss GenAI in closed communities or offline settings. Although care was taken in cleaning and annotating the data, sentiment classification is still a probabilistic process. Misclassification will always be a part of the process, especially in cases of highly ambiguous or sarcastic posts, even if a strong model is used. Since keyword-based filters have been utilized for education and healthcare, it implies that some very relevant posts that use less typical language might not have been detected, and a small number of posts that are only loosely related may have beenincluded. Lastly, the study's scope is relatively narrow, spanning from November 2022 to December 2023. While this offers a more detailed temporal view than many early studies, it still does not capture longer-term shifts that may emerge as GenAI becomes more deeply embedded in everyday systems and as regulatory responses continue to develop.

6.2. Directions for Future Research

Future research could extend this framework in several directions to capture richer and more nuanced public perceptions of GenAI. First, multimodal SA incorporating textual, visual, and audiovisual content—such as images and videos—could be employed to better interpret sentiment expressed through memes, screenshots, and short-form media. Second, fine-grained emotion classification may complement polarity-based SA by distinguishing affective responses such as trust, fear, enthusiasm, or anxiety toward GenAI systems. Third, stance detection techniques could be integrated to identify whether users support, oppose, or remain neutral toward the adoption of GenAI, particularly in sensitive domains such as education and healthcare. Finally, extending the analysis across multiple social media platforms would enable comparative insights into how platform-specific affordances shape public discourse and engagement with GenAI technologies.

References

Abdallah, N., Katmah, R., Khalaf, K. & Jelinek, H.F. (2025) 'Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration', *Social Sciences and Humanities Open*, 12, 101866. Available at: https://doi.org/10.1016/j.ssaho.2025.101866.

Ainapure, B.S. et al. (2023) 'Sentiment analysis of COVID-19 tweets using deep learning and lexicon-based approaches', *Sustainability*, 15, 2573. Available at: https://doi.org/10.3390/su15032573.

Ali, K., Barhom, N., Tamimi, F. & Duggal, M. (2024) 'ChatGPT—A double-edged sword for healthcare education? Implications for assessments of dental students', *European Journal of Dental Education*, 28, pp. 206–211. Available at: https://doi.org/10.1111/eje.12937.

Alslaity, A. & Orji, R. (2024) 'Machine learning techniques for emotion detection and sentiment analysis: Current state, challenges, and future directions', *Behaviour & Information Technology*, 43(1), pp. 139–164. Available at: https://doi.org/10.1080/0144929X.2022.2156387.

Alsemaree, O., Alam, A.S., Gill, S. & Uhlig, S. (2024) 'Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions', *Heliyon*, 10(9), e27863. Available at: https://doi.org/10.1016/j.heliyon.2024.e27863.

Baxter, P., Li, M-H., Wei, J. & Koizumi, N. (2025) 'Public versus academic discourse on ChatGPT in health care: Mixed methods study', *JMIR Infodemiology*, 5(1). Available at: https://doi.org/10.2196/64509.

Bermudez-Sosa, H.J., Olarte-Henao, J. & Rojas-Berrío, S.P. (2025) 'Sentiment and emotion analysis from textual information: A systematic literature review', *Journal of Information Science*. Available at: https://doi.org/10.1177/01655515251353170.

Chan, C.K.Y. & Hu, W. (2023) 'Students' voices on generative AI: Perceptions, benefits, and challenges in higher education', *International Journal of Educational Technology in Higher Education*, 20(1), pp. 1–18. Available at: https://doi.org/10.1186/s41239-023-00411-8.

Chomiak-Orsa, I. (2025) 'Sentiment and emotion analysis of public discourse on ChatGPT using VADER sentiment analysis', *Journal of Digital Society*, 1(1). Available at: https://doi.org/10.63913/jds.v1i1.1.

Darraz, N. et al. (2024) 'Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems', *Expert Systems with Applications*, 261, 125533. Available at: https://doi.org/10.1016/j.eswa.2024.125533

Demirel, S., Kahraman-Gokalp, E. & Gündüz, U. (2024) 'From optimism to concern: Unveiling sentiments and perceptions surrounding ChatGPT on Twitter', *International Journal of Human–Computer Interaction*. Available at: https://doi.org/10.1080/10447318.2024.2392964.

Diri, G., Obiorah, P. & Du, H. (2025) 'Comparative study of sentiment analysis techniques: Traditional machine learning vs deep learning approaches', in *InSITE 2025: Informing Science + IT Education Conferences*. Hiroshima: Informing Science Institute. Available at: https://doi.org/10.28945/5490.

Dos, I. (2025) 'A systematic review of research on ChatGPT in higher education', *The European Educational Researcher*, 8(2), pp. 59–76. Available at: https://doi.org/10.31757/euer.824.

Iqbal, U. et al. (2025) 'Impact of large language model (ChatGPT) in healthcare: An umbrella review and evidence synthesis', *Journal of Biomedical Science*, 32(1), 45. Available at: https://doi.org/10.1186/s12929-025-01131-z.

Jahanbin, K. & Ali, M. (2023) 'Aspect-based sentiment analysis of Twitter influencers to predict cryptocurrency trends using hybrid deep transfer learning models', *IEEE Access*, pp. 1–1. Available at: https://doi.org/10.1109/access.2023.3327060.

Jin, Y. & Sercu, L. (2025) 'ChatGPT interventions in higher education: A systematic review of experimental studies', *Journal of Computer Assisted Learning*, 41(4). Available at: https://doi.org/10.1111/jcal.70072.

Khan, N. et al. (2024) 'Global insights and the impact of generative AI-ChatGPT on multidisciplinary fields: A systematic review and bibliometric analysis', *Connection Science*, 36(1). Available at: https://doi.org/10.1080/09540091.2024.2353630.

Koonchanok, R., Pan, Y. & Jang, H. (2024) 'Public attitudes toward ChatGPT on Twitter: Sentiments, topics, and occupations', *Social Network Analysis and Mining*, 14(1). Available at: https://doi.org/10.1007/s13278-024-01260-7.

Krishna, K.M. et al. (2023) 'A CNN-LSTM-based hybrid deep learning approach for sentiment analysis on Monkeypox tweets', *New Generation Computing*. Available at: https://doi.org/10.1007/s00354-023-00227-0.

Lee, J., Kim, H. & Kron, F. (2024) 'Virtual education strategies in sustainable healthcare and medical education: A topic modelling analysis of four decades of research', *Medical Education*, 58, pp. 47–62. Available at: https://doi.org/10.1111/medu.15202.

Li, L. et al. (2022) 'Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews', *Computers & Education*, 176, 104354. Available at: https://doi.org/10.1016/j.compedu.2021.104354.

Lian, Y. et al. (2024) 'Public attitudes and sentiments toward ChatGPT in China: A text mining analysis based on social media', *Technology in Society*, 76, 102442. Available at: https://doi.org/10.1016/j.techsoc.2023.102442.

Liu, Y. & Lyu, Z. (2024) 'Changes in public perception of ChatGPT: A text mining perspective based on social media', *International Journal of Human–Computer Interaction*. Available at: https://doi.org/10.1080/10447318.2024.2406966.

Madsen, D.Ø. & Toston, D.M. II (2025) 'ChatGPT and digital transformation: A narrative review of its role in health, education, and the economy', *Digital*, 5(3), 24. Available at: https://doi.org/10.3390/digital5030024.

Mao, Y., Liu, Q. & Zhang, Y. (2024) 'Sentiment analysis methods, applications, and challenges: A systematic literature review', *Journal of King Saud University – Computer and Information Sciences*, 36(4), 102048. Available at: https://doi.org/10.1016/j.jksuci.2024.102048.

Miyazaki, K. et al. (2024) 'Public perception of generative AI on Twitter: An empirical study based on occupation and usage', *EPJ Data Science*, 13(2). Available at: https://doi.org/10.1140/epjds/s13688-023-00445-y.

Munaye, Y.Y. et al. (2025) 'ChatGPT in education: A systematic review on opportunities, challenges, and future directions', *Algorithms*, 18(6), 352. Available at: https://doi.org/10.3390/a18060352.

Naing, S. & Udomwong, P. (2024) 'Public opinions on ChatGPT: An analysis of Reddit discussions using sentiment analysis, topic modelling, and SWOT analysis', *Data Intelligence*, pp. 1–50. Available at: https://doi.org/10.1162/dint_a_00250.

Qi, Y. & Shabrina, Z. (2023) 'Sentiment analysis using Twitter data: A comparative application of lexicon- and machine-learning-based approaches', *Social Network Analysis and Mining*, 13(1). Available at: https://doi.org/10.1007/s13278-023-01030-x.

Raees, M. & Fazilat, S. (2024) 'Lexicon-based sentiment analysis on text polarities with evaluation of classification models', *arXiv*. Available at: https://doi.org/10.48550/arxiv.2409.12840.

Ramezani, E.B. (2025) 'Sentiment analysis applications using deep learning advancements in social networks: A systematic review', *Neurocomputing*, 634, 129862. Available at: https://doi.org/10.1016/j.neucom.2025.129862.

Rodríguez-Ibánez, M. et al. (2023) 'A review on sentiment analysis from social media platforms', *Expert Systems with Applications*, 223, 119862. Available at: https://doi.org/10.1016/j.eswa.2023.119862.

Sallam, M. (2023) 'ChatGPT utility in healthcare education, research, and practice: Systematic review on promising perspectives and valid concerns', *Healthcare*, 11, 887. Available at: https://doi.org/10.3390/healthcare11060887.

Shaik, T. et al. (2023) 'Sentiment analysis and opinion mining on educational data: A survey', *Natural Language Processing Journal*, 2, 100003. Available at: https://doi.org/10.1016/j.nlp.2022.100003.

Shimaoka, A.M., Ferreira, R.C. & Goldman, A. (2024) 'The evolution of CRISP-DM for data science: Methods, processes and frameworks', *SBC Reviews on Computer Science*, 4(1), pp. 28–43. Available at: https://doi.org/10.13140/RG.2.2.22493.42721.

Su, Y. & Kabala, Z.J. (2023) 'Public perception of ChatGPT and transfer learning for tweets sentiment analysis using Wolfram Mathematica', *Data*, 8(12), 180. Available at: https://doi.org/10.3390/data8120180.

Tabany, M. & Gueffal, M. (2024) 'Sentiment analysis and fake Amazon reviews classification using SVM supervised machine learning model', *Journal of Advances in Information Technology*, 15(1), pp. 49–58. Available at: https://doi.org/10.12720/jait.15.1.49-58.

Tan, K.L., Lee, C.P. & Lim, K.M. (2023) 'A survey of sentiment analysis: Approaches, datasets, and future research', *Applied Sciences*, 13(7), 4550. Available at: https://doi.org/10.3390/app13074550.

Tao, Y. & Shen, Q. (2025) 'Academic discourse on ChatGPT in social sciences: A topic modelling and sentiment analysis of research article abstracts', *PLOS ONE*, 20(10). Available at: https://doi.org/10.1371/journal.pone.0334331.

Williams, L. & Burnap, P. (2025) 'The emotional landscape of technological innovation: A data-driven case study of ChatGPT's launch', *Informatics*, 12(3), 58. Available at: https://doi.org/10.3390/informatics12030058.

Wu, S. & Gao, Y. (2024) 'Machine learning approach to analyse the sentiment of airline passengers' tweets', *Transportation Research Record*, 2678(2), pp. 48–56. Available at: https://doi.org/10.1177/03611981231172948.

Zhang, P. & Kamel Boulos, M.N. (2023) 'Generative AI in medicine and healthcare: Promises, opportunities and challenges', *Future Internet*, 15(9), 286. Available at: https://doi.org/10.3390/fi15090286.